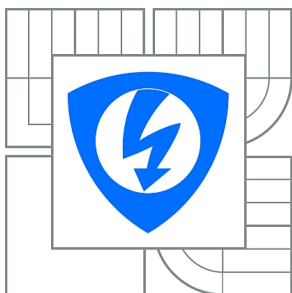




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNologií
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

PRAVDĚPODOBNOSTNÍ MODEL SÍŤOVÉHO PROVOZU

PROBABILITY-BASED MODEL FOR NETWORK TRAFFIC

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

PETER BEDNÁŘ

VEDOUCÍ PRÁCE
SUPERVISOR

doc. Ing. KAROL MOLNÁŘ, Ph.D.

BRNO 2010



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Peter Bednár
Ročník: 3

ID: 78059
Akademický rok: 2009/2010

NÁZEV TÉMATU:

Pravidelnostní model síťového provozu

POKYNY PRO VYPRACOVÁNÍ:

Identifikujte základní typy provozu v IP síti a definujte charakteristické rysy jednotlivých typů. Seznamte se se stávajícími modely běžných síťových služeb. V rámci bakalářské práce se zaměřte především na služby peer-to-peer, HTTP a VoIP. Identifikujte parametry jednotlivých modelů a navrhnete univerzální model, který bude schopen modelovat chování těchto služeb. Při návrhu zohledněte požadavek na rozšiřitelnost pro další typy aplikací. Navrhnete metodu nalezení potřebných parametrů univerzálního modelu pomocí analýzy reálného síťového provozu. Proveďte softwarovou realizaci daného modelu.

DOPORUČENÁ LITERATURA:

- [1] STAEHLE, D., LEIBNITZ, K., TRAN-GIA, P. Source Traffic Modeling of Wireless Applications, Universität Würzburg, Institut für Informatik, Würzburg, 2000
- [2] KLEMM, A., LINDEMANN, CH., LOHMANN, M. Traffic Modeling and Characterization for UMTS Networks, Proc. of the Globecom, Internet Performance Symposium, San Antonio, pp. 1741-1746, 2001.

Termín zadání: 29.1.2010

Termín odevzdání: 2.6.2010

Vedoucí práce: doc. Ing. Karol Molnár, Ph.D.

prof. Ing. Kamil Vrba, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva těchto osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení druhé části, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Anotácia

Táto práca identifikuje základné druhy prenosu v IP sieťach a opisuje ich hlavné črty. Zaoberá sa základnými sieťovými službami HTTP, VoIP a elektronickej pošty, pričom sa sústreďuje na ich základné pravdepodobnostné modely v sieťovej premávke, ktoré aplikuje pri softvérovej simulácii reálnej siete.

Kľúčové slová

pravdepodobnosť, paket, fronta, model, prenos, protokol, sieť, užívateľ, aktívna fáza, neaktívna fáza, hlavný objekt, vedľajší objekt, tok dát smerom k užívateľovi, tok dát smerom od užívateľa, internetový prehliadač, nečinnosť, proces, interval, rozdiel, spojenie, sťahovať, načítavať, požiadavka, stránka, záznam, obsah, správanie

Annotation

This work identifies several basic types of traffic in IP networks and describes main features. The basic network services such as HTTP, VoIP and electronic mail are discussed focusing mainly on HTTP and VoIP probability-based models for network traffic which are applied to the software simulation of real network.

Key words

probability, packet, front, model, traffic, protocol, network, user, ON phase, OFF phase, main object, inline object, downstream, upstream, web browser, idle, process, period, difference, connection, load, download, request, page, log, content, behaviour

BEDNÁŘ, P. *Pravděpodobnostní model síťového provozu*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2010. 50 s. Vedoucí bakalářské práce doc. Ing. Karol Molnár, Ph.D.

Prehlásenie o pôvodnosti práce

Prehlasujem, že svoju bakalársku prácu na tému *Pravdepodobnostný model sieťovej premávky* som vypracoval samostatne pod vedením vedúceho bakalárskej práce s použitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú uvedené v zozname literatúry na konci práce.

Ako autor uvedenej bakalárskej práce ďalej prehlasujem, že v súvislosti s vytvorením tejto diplomovej práce som neporušil autorské práva tretích osôb a nezasiahol som nedovoleným spôsobom do cudzích autorských práv osobnostných a som si plno vedomý následkov porušenia ustanovení §11 a nasledujúceho autorského zákona č. 121/2000 Sb., vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovenia č.152 trestného zákona č. 140/1961 Sb.

V Brne dňa

.....
podpis autora

PodĎakovanie

Ďakujem vedúcemu bakalárskej práce Doc. Ing. Karolovi Molnárovi, Ph.D., za užitočnú metodickú pomoc a cenné rady pri spracovaní práce.

V Brne dňa

.....
podpis autora

OBSAH

Zoznam obrázkov	7
Zoznam tabuliek	7
1. Úvod	8
1.1 Model sieťovej premávky	8
1.2 Postup návrhu sieťového prenosu	8
2. Základné druhy prenosov v sieti	9
2.1 Základný model hlasových sietí	9
2.2 Základný model dátových sietí	10
3. Základné modely sieťového prenosu najčastejšie používaných aplikácií ...	11
3.1 Hlavné črty	11
3.2 Návrh predikcie pomeru využívania jednotlivých služieb užívateľmi pri prechode na rýchlejšie pripojenie ADSL	12
3.3 Základné modely sieťového prenosu	14
4. Model webového prenosu HTTP	14
4.1 Súhrn existujúcich webových modelov	15
4.2 Model správania	17
4.3 Navrhovaný model	19
4.4 Superpozícia dátových tokov viacerých užívateľov	21
4.5 Model prenosu jedného užívateľa	22
4.6 Model HTTP	22
5. Model prenosu hlasu a videa VoIP	26
5.1 Základný model komunikácie dvoch užívateľov	27
5.2 VoIP protokol - architektúra	27
5.3 Sieťová premávka	29
5.4 Model sieťovej premávky VoIP	29
5.5 Model elektronickej pošty	34
6. Softwarová simulácia modelu prenosu	35
6.1 Zostava reálnej počítačovej siete	35
6.2 OPNET IT Guru Academic Edition 9.1	36
6.3 Postup vytvorenia referenčného modelu siete	37
6.4 Simulácia modelu siete	39
6.5 Analýza prenosov v reálnej sieti	40
6.6 Simulovaná a reálna sieťová premávka	41
7. Záver	42
8. Dodatok	45
8.1 Vybrané pojmy z teórie pravdepodobnosti a štatistiky	45
Zoznam použitej literatúry	51
Zoznam použitých skratiek	53
Zoznam príloh	54

Zoznam obrázkov

1. Príklad odchytenej komunikácie protokolu HTTP
2. Model správania sa užívateľa
3. Znázornenie typu dátových tokov generovaných DSL užívateľom
4. Znázornenie superpozície dátových tokov prichádzajúcich do centrálneho uzla siete poskytovateľa
5. Model správania sa jedného DSL užívateľa
6. Výsledná HTTP premávka na hlavnom obslužnom zariadení (router ISP)
7. Proces aktivity pre www
8. Model komunikácie 2 užívateľov prostredníctvom VoIP
9. Pravdepodobnosti časov prichádzajúcich hovorov
10. Pravdepodobnosti časov držania hovorov
11. Pravdepodobnosti períód ON/OFF
12. Pravdepodobnosti periódy ON pre kodek G.729
13. Pravdepodobnosti periódy OFF pre kodek G.729
14. Distribučné funkcie veľkostí E-mailov
15. Topológia reálnej siete
16. Zapojenie pre odchytyvanie paketov
17. Prenosy aplikačných protokolov HTTP, SMTP a POP3
18. Menu programu IT Guru
19. Menu Project manager
20. Paleta prvkov knižnice ethernet
21. Zostavený model simulovanej siete v prostredí IT Guru
22. Scenár referenčného modelu v prostredí IT Guru
23. Definícia aplikácií s nastaveniami pravdepodobnosti prenosu
24. Voľba štatistík a výstupný graf simulácie
25. Zastúpenie protokolov v odchytenej premávke
26. Grafický výstup zo simulovanej a reálnej webovej premávky
27. Grafický výstup zo simulovanej a reálnej premávky elektronickej pošty

Zoznam tabuliek

1. Hodnoty parametrov pre model HTTP
2. Signalizačné a transportné protokoly VoIP
3. Hodnoty zistených veľkostí E-mailov
4. Štatistika súboru odchytenej premávky.

1. Úvod

Tradícia v tvorbe telekomunikačných sietí mala v minulosti malé alebo žiadne základy opierajúce sa o analýzy alebo architektonický vývoj modelov. Sieťový dizajnéri sa často spoliehali na technológie, ktoré im boli známe z nedávnej minulosti alebo nové, ktoré im odporúčali výrobcovia zariadení, prípadne ich konzultanti. Tento prístup sa s rozvíjajúcimi s požiadavkami na telekomunikačné siete začal javiť ako nevhodný a spôsoboval značné problémy. Pri návrhu bolo treba postupovať podľa požiadaviek užívateľov a rozvíjajúcich sa trendov v oblasti telekomunikácií.

1.1 Model sieťovej premávky

Vybranie vhodného zdrojového modelu sieťového prenosu zvyčajne odráža správanie sa užívateľov v telekomunikačnom systéme. Je to veľmi dôležitý krok pre vytvorenie správne pracujúcich telekomunikačných sietí, najmä sietí nových generácií, ktoré musia svojim užívateľom flexibilne poskytovať mnohé druhy špecifických služieb, napríklad VoIP, ktoré vyžaduje detailný popis modelu sieťovej premávky. Na rozdiel od existujúceho GSM systému, ktorý primárne obsluhuje len hlasových užívateľov a zriedka len sekundárne poskytuje dátové služby, napríklad prostredníctvom GPRS, existujú aj dátové siete, ktoré musia rovnako spoľahlivo prenášať rôzne druhy prenosu súčasne, pričom každý prenášaný paket môže mať inú prioritu. Bez správneho zhromaždenia požiadaviek a analýz si preto nemôžeme byť istí, či nami vytvorený model siete bude správne uspokojovať požiadavky svojich koncových užívateľov. Z tohto dôvodu sú analýza siete, jej architektúra a tvar fundamentálne zložky pre rozvoj telekomunikačných sietí.

1.2 Postup návrhu sieťového prenosu

Sieťovú analýzu, architektúru a tvar môžeme porovnávať taktiež k iným inžinierskym postupom a pristupovať k ich riešeniu nasledujúcimi, podobnými krokmi :

- Definovanie riešeného problému
- Zoskupenie požiadaviek užívateľov
- Monitoring súčasnej siete, jej modelu a prostredia
- Analýza dát
- Návrh niekoľkých možností pre riešenie problému
- Zhodnotenie a optimalizácia návrhov riešení
- Výber jedného, alebo niekoľkých z návrhov
- Plánovanie implementácie, realizácia

Základ tvorby návrhu každého sieťového modelu by mal byť odhad všetkých očakávaní užívateľa a patričné prihliadanie k nim. Táto myšlienka nemá nabádať k tomu, aby sa užívateľom dávali falošné očakávania ani aby sme užívateľom sľubovali nereálne skutočnosti, pretože by to mohlo viesť k problémom neskôr pri realizácii projektov. Cieľom pri návrhu je poskytnúť presný a reálny pohľad na technické problémy a ukázať cestu, ktorou sa tieto problémy dajú vyriešiť.

2. Základné druhy prenosov v sieti

Základné druhy prenosu poznáme už z čias telegrafu a telefónu. V najzákladnejšom ponímaní ich môžeme rozdeliť na hlasový a dátový prenos informácií. Z pohľadu typov sieťových spojení, akými sú tieto dáta prenášané cez komunikačné kanály, na dva základné druhy a to siete s komutáciou okruhov (so spájaním okruhov) a komutáciou paketov (paketový prenos).

V sieťach s komutáciou okruhov ide zväčša o synchronný prenos s vytvorením spojenia po celú dobu prenosu. Príkladom je synchronný prenosový mód STM (synchronous transfer mode) pre prenos dát. Výhody týchto sietí sú spojené s nízkym oneskorením prenosu informácií a sú vhodné väčšinou pre „real-time“ prenosi v PSTN sieťach (prenos hlasu). Pre dátové prenosi sú však nedostačujúce najmä pre obmedzené prenosové rýchlosti. Z týchto dôvodov boli pre dátový prenos vyvinuté siete s komutáciou paketov úzko spojené s využitím protokolu IP. Je pre ne charakteristický asynchronný druh prenosu (ATM, asynchronous transfer mode) pre prenos krátkych správ ako aj väčších dátových objemov. Nevýhodou týchto sietí bolo väčšie zdržanie paketov v sieti, menej vhodné pre „real-time“ prenosi. Túto nevýhodu sa však postupom času s príchodom novších protokolov (RTP, RTCP) a kvality služieb QoS podarilo do určitej miery eliminovať a prispôsobiť tak tieto siete aj na prenos hlasu a videa prostredníctvom VoIP (Voice over Internet Protocol) a ďalších protokolov.

2.1 Základný model hlasových sietí

Podobne ako v dnešných sofistikovaných modeloch opísaných v literatúre môžeme charakterizovať hlasových užívateľov do vrstvenej štruktúry. V najširšom ponímaní je prenos charakteristický aktivitou užívateľa zahŕňajúcu čas zostavenia spojenia až po jeho ukončenie. Parametre pre tento typ prenosu môžeme získať z empirických meraní. Pozostávajú z časov medzi zostavením spojenia a celkových časov spojenia. V konvenčných systémoch a klasických telefónnych sieťach bolo akceptované exponenciálne rozdelenie pre oba parametre, ktoré vedie k modelu s Poissonovým rozdelením pravdepodobnosti opisu pre hlasových užívateľov [1].

Pri porovnaní prenosov hlasových a dátových užívateľov zisťujeme, že prenosy hlasových užívateľov sú veľmi jednotvárne. Prenos generovaný dátovými užívateľmi veľmi závisí od druhu aplikácie. Rozdiely medzi časmi prichádzajúcich paketov ako aj ich celková dĺžka môžu byť veľké. Z tohto dôvodu sa ukazuje, že Poissonov model nie je pre dátové siete vyhovujúci.

2.2 Základný model dátových sietí

Základný model dátového prenosu v prostredí siete ISDN je v [6] odvodený z procesov prichádzajúcich paketov od jedného užívateľa. Tento model zachytáva koreláciu v distribúcii medzi-prichádzajúcich časov paketových tokov (doby medzi susednými paketovými tokmi). Táto štúdia ďalej ukazuje, že aplikácie, ktoré vyžadujú interakcie (čaká sa na reakciu od užívateľa) majú správanie LRD (long range dependence) a sú v dobrej zhode s Paretovým rozdelením pravdepodobnosti (Definícia 8.13, 8.14). Jeho použitie však môže byť problematické nakoľko priemer a štandardná odchýlka môžu pre určité prípady nadobúdať nekonečných hodnôt.

Pre internetový prenos boli v minulosti odvodené rôzne modely analýz protokolov telnet, NNTP, SMTP a FTP založené na predchádzajúcich pozorovaniach uvedených v [8]. Bolo zistené, že štatistické modelovanie prenosu pre siete WAN je veľmi náročné. Môžeme však postupovať pomocou vytvorenia jednoduchých analytických metód, ktoré sa presnosťou blížia k dobrým odhadom. Príklad takéhoto postupu je uvedený v [9], kde je ukázané, že prichádzajúce TCP spojenie iniciované užívateľom (napríklad vzdialený prístup, prenos súborov) je možné modelovať Poissonovým rozdelením, avšak intervaly pre medzi-prichádzajúce pakety každého spojenia vykazujú prudké zmeny (burstiness). Vede to k nesprávnemu odhadu výkonnostných parametrov ako sú napríklad priemerné oneskorenia paketov alebo maximálna veľkosť frontu.

V minulých štúdiách bol tiež zistený podobný vzťah pre LAN a WAN siete. Dátové súbory s rozdelením pravdepodobnosti s tzv. ťažkými chvostmi a operácie spoľahlivého TCP prenosu spolu s jeho kontrolným mechanizmom toku dát majú štruktúru LRD (long range dependency). Pre porovnanie nekontrolované dátové toky a nespoľahlivé UDP prenosy vykazovali veľmi malú závislosť LRD. Tiež bolo uvedené, že výkonnosť v zmysle oneskorenia dátových front bola ovplyvňovaná samopodobnosťou (v zmysle teórie chaosu) pričom výkonnostné parametre ovplyvňujúce priepustnosť ako strata paketu a spätný prenosový pomer zostali väčšinou neovplyvnené pri použití spoľahlivého prenosového protokolu s riadením toku.

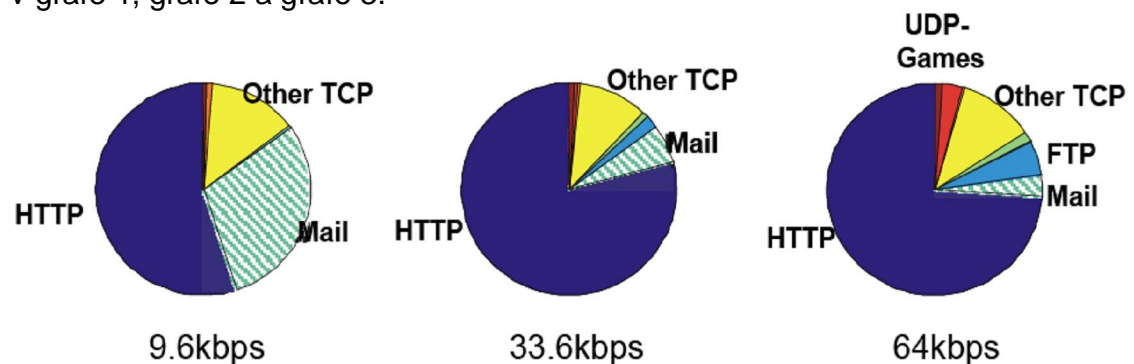
Odhliadnuc od konkrétnej aplikácie je správanie užívateľa do istej miery ovplyvnené aj inými externými vplyvmi ako napríklad tarifikačia. Ak je služba

z pohľadu užívateľa drahá, menej zákazníkov ju bude pochopiteľne využívať, alebo skrátiť dobu jej využitia. Navyše v dôsledku adaptačného mechanizmu TCP prenosové objemy závisia na záťaži siete a aj od spracovateľských schopnostiach osloveného servera. Znamená to, že správanie užívateľa nie je určená iba použitou aplikáciou, ale aj na ťažšie predvídateľných vplyvoch.

3. Základné modely sieťového prenosu najčastejšie používaných aplikácií

3.1 Hlavné črty

V predchádzajúcej časti sme sa pokúsili charakterizovať TCP ako jeden spoločný prúd dát slúžiaci pre viaceré aplikácie súčasne. Avšak, keď sa pozrieme na rozličné aplikácie ako sú napríklad WWW, E-mail, FTP zistíme, že tieto generujú dátové toky, ktoré sú vo svojich charakteristických črtách navzájom odlišné. V [13] bolo pozorované správanie užívateľov, ktorí sa pripájali do siete pomocou vytáčaného pripojenia s rôznou prenosovou rýchlosťou pripojenia ich liniek. Výsledky jeho pozorovaní môžeme vidieť v grafe 1, grafe 2 a grafe 3.



V poradí zľava, graf 1, graf 2 a graf 3 vyjadrenie pomeru využívania jednotlivých služieb užívateľmi s rôznymi prenosovými rýchlosťami [4].

Z grafu vyplýva, s rastúcou rýchlosťou pripojenia rastie zložka pomerového zastúpenia prenosu HTTP a klesá zložka zastúpenia mailových služieb. Rastie aj zložka UDP Games, čo sa dá vysvetliť vyššou prenosovou rýchlosťou pripojenia, nakoľko pomalšie pripojenia požiadavkám rýchleho prenosu dát potrebného pre hranie on-line hier nevyhovuje. Na základe poznatkov z takéhoto druhu výskumu je teraz možné pokúsiť sa odhadnúť správanie užívateľov pri prechode z vytáčaného pripojenia na pripojenie xDSL, u nás zastúpené technológiou ADSL.

3.2 Návrh predikcie pomeru využívania jednotlivých služieb užívateľmi pri prechode na rýchlejšie pripojenie ADSL

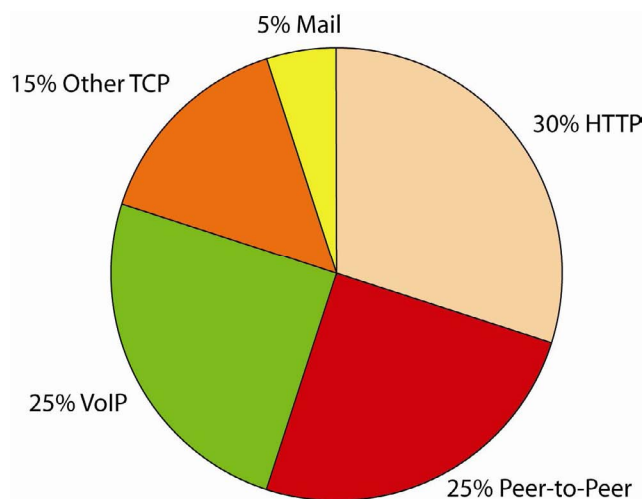
Pre predpoveď budúceho zloženia sieťového prenosu užívateľmi musíme zvážiť niekoľko aspektov.

1. ADSL užívatelia budú mať k dispozícii väčšiu prenosovú rýchlosť, čo bude mať za následok väčšie nároky na hardwarové zložky siete. Prenosová rýchlosť sa v dnešnej dobe pohybuje typicky na úrovni 2Mbps downstream a 512Kbps upstream (december 2008).
2. Od doby vytáčaných pripojení sa zmenil aj obsah WWW stránok. K jednoduchým formám obrázkov a textu pribudol ďalší na dátový objem náročnejší multimediálny obsah, reprezentovaný najmä pohyblivými grafickými prvkami flash, náročnejším grafickým spracovaním, zvukom a prehrávaním video súborov v prehrávačoch implementovaných priamo v okne web prehliadača.
3. Rozšírenie internetovej telefónie VoIP (Voice over Internet Protocol), ktorá vyžaduje stabilnú rýchlosť, stabilitu a nízke zdržanie dátových jednotiek v sieti.
4. Užívatelia, ktorí nebudú dátovo obmedzovaní budú zaťažovať sieťové prostriedky využitím dátových výmenných sietí typu peer-to-peer. Predpoklad správania sa užívateľa s dátovým obmedzením je využívanie služieb HTTP, prípadne ich multimediálneho obsahu a E-mailovej komunikácie.
5. Užívateľské aplikácie budú väčšinou rovnaké pre užívateľa pracujúceho napríklad na stolnom počítači ako aj na prenosnom notebooku.
6. Účtovanie poplatkov v ADSL nebude závislé na dĺžke trvania spojenia, ale typicky na množstve prenesených údajov (neplatí pre užívateľov s dátovo neobmedzeným pripojením).

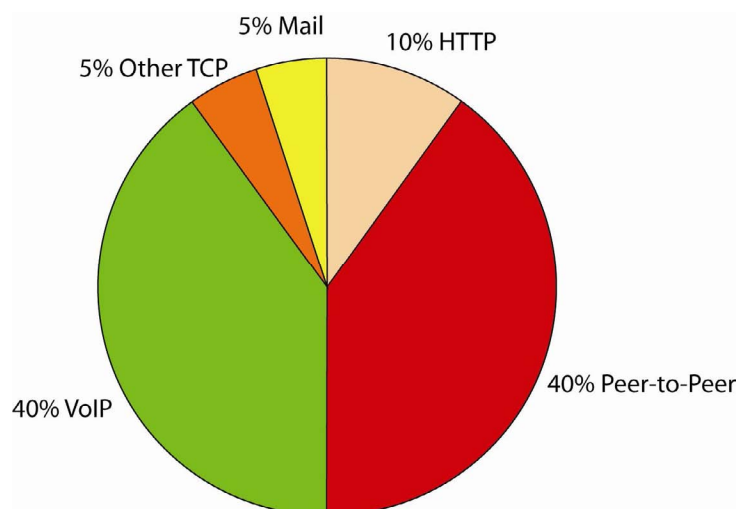
Pri návrhu modelu sieťového prenosu v ADSL sa musí prihliadať na všetky tieto body. Odhaduje sa, že situácia pomerného zastúpenia jednotlivých služieb v modeli sieťového prenosu ADSL bude značne odlišná od technológie vytáčaného pripojenia. Nakoľko nie sú k dispozícii aktuálnejšie údaje získané oficiálnym výskumom, sú použité informácie získané pozorovaním užívateľov domácnosti s aktívnym ADSL pripojením, z time logov od poskytovateľa služieb a programovým meraním (Observer, WireShark). Pre názornosť je uvedený Graf 4 dátovo obmedzeného užívateľa a Graf 5 dátovo neobmedzeného užívateľa a ich predpokladaného zloženia sieťového

prenosu v ADSL. Z dôvodu väčšej šírky pásma je tu predpoklad reálneho využitia služieb VoIP.

Z Grafu 4 je zrejmé, že väčšinu zo všetkých prenosov, až 80%, by mali tvoriť tri podstatné zložky, ktorými sú prehliadanie WWW stránok s multimediálnym obsahom, využívanie výmenných sietí a internetová telefónia. Pozorovanie vychádza z faktu, že užívateľovi je za určité obdobie pridelený istý voľný objem dát, po ktorom prečerpaní sú dáta spoplatňované vysokými tarifikami. Časť označená oranžovou farbou zahŕňa rôzne iné služby napríklad prenosy on-line hier, telnet, atď.



Graf 4. Odhad zloženia pomerového zastúpenia typu dátových prenosov v modeli sieťového prenosu dátovo obmedzeného užívateľa.



Graf 5. Odhad zloženia pomerového zastúpenia typu dátových prenosov v modeli sieťového prenosu dátovo neobmedzeného užívateľa.

Z Grafu 5 je zrejmé, že väčšinu zo všetkých prenosov, až 80% by mali tvoriť dve podstatné zložky, ktorými sú využívanie výmenných sietí a internetová telefónia. Pozorovanie vychádza z faktu, že užívateľ nie je časovo ani dátovo obmedzovaný a za rovnaké obdobie prenesie väčší objem dát ako užívateľ so spomínaným obmedzením. Časť označená oranžovou farbou opäť zahŕňa rôzne iné služby napríklad prenosy on-line hier, telnet, atď. V tomto prípade klesá zložka percentuálneho zastúpenia HTTP prenosov s multimediálnym obsahom. Tieto údaje boli realizované štatisticky, kde každý respondent odpovedal v snahe s čo najväčšou presnosťou o službách ktoré najviac využíva a ich približnom pomerovom zastúpení. Je zrejmé., že takto zistené údaje sú veľmi nepresné a nebudeme z nich ďalej už vychádzať. Pri analýze budeme vychádzať z reálne odchytených dát programom WireShark.

3.3 Základné modely sieťového prenosu

Druhy základných modelov prenosu pozostávajú z dvoch častí: z prichádzajúcich procesov pre užívateľské aktivity a z procesov opisujúcich fázy ich činnosti (aktivity). Modely pre jednotlivé fázy činností týchto služieb budú opísané jednotlivo. Prichádzajúci proces určuje prípad, keď užívateľ začína svoju aktivitu. Tieto prípady sú označované ako prichádzajúce časy. Okamih v ktorom užívateľ začína v prípade HTTP samotné prezeranie webu je označovaný ako prichádzajúci čas HTTP aplikácie. Prichádzajúci čas môže byť opísaný buď ako medzi-prichádzajúci čas, t.j. ako čas medzi dvoma nasledujúcimi príchodmi alebo ako distribúcia fázy OFF, t.j. čas medzi koncom aktivity užívateľa a začiatkom jeho ďalšej aktívnej fázy. Prístup použitia medzi-prichádzajúcich časov je vhodný ak pripúšťame naraz väčšie množstvo užívateľov a dĺžky aktívnych fáz vytvárajú len zlomok dĺžky fáz OFF. Najčastejšie vyžívaný prichádzajúci proces tohto typu je Poissonov proces s negatívne exponenciálne rozloženými časovými intervalmi. Takto podobne môžeme postupovať pri sledovaní väčšiny služieb snád' s výnimkou VoIP (resp. XoIP), kde je komunikácia zložitejšia. Cieľ je zložením parciálnych modelov komunikácií vytvoriť model celkového pravdepodobnostného modelu sieťovej premávky, ktorý odzrkadľuje konkrétnu sieť, prípadne typ siete.

4. Model webového prenosu HTTP

Webový prenos je v dnešnej dobe najdôležitejšia forma aplikácie využívaná Internetovou komunitou. Termín webový prenos zahŕňa všetky HTTP prenosy generované počas komunikácie typickým webovým prehliadačom, akým sú napríklad Mozilla, Firefox alebo Internet Explorer. Nie všetky prenosy prebiehajú však rovnako. Sú tu diferencie či už v rozdielnych verziách HTTP alebo medzi prehliadačmi. Tieto odlišné verzie ovplyvňujú veľaokrát spôsob akým sú webové stránky sťahované.

Aby sme dokázali pochopiť odlišné metódy prenášania stránok, treba si pripomenúť ich základnú štruktúru. Webová stránka sa skladá z ASCII textu, kódu HTML. Časti stránok odkazujú na hlavné objekty. V kóde HTML sa obrázky, video alebo objekty ako Java alebo Flash vkladajú ako odkaz na oddelené súbory zvlášť uložené na serveri označované ako inline objekty. Sťahujúca procedúra WWW stránky začína, keď užívateľ klikne na odkaz odkazujúci na túto stránku. Následne je zaslaný príkaz *getrequest* do URL (Universal Resource Locator). K prenosu týchto požiadaviek sa nadviaže TCP spojenie medzi užívateľom a web serverom. Po prijatí požiadavky pošle web server súbory prehliadaču. Ten vykoná rozbor HTML kódu a použitie inline objektov, ktoré ešte neboli doteraz prenesené.

V tomto bode sa začínajú jednotlivé verzie od seba navzájom líšiť. Prvou a najmenej efektívnou metódou je zostavenie samostatného TCP spojenia pre každý inline objekt. Neskoršie verzie zlepšujú účinnosť tejto metódy sťahovaním niekoľkých inline objektov v jednom TCP spojení a zoskupením niekoľkých TCP spojení naraz paralelne. Niekedy sa využíva aj možnosť podržania aktívneho TCP spojenia aj po prenesení celého obsahu stránky. Toto riešenie je výhodné, ak aj ďalšia požadovaná stránka bude uložená na tom istom serveri. V tom prípade nebude treba otvárať nové TCP spojenie a využije sa existujúce. V dôsledku pomalého štartovacieho mechanizmu TCP dosiahne sa týmto spôsobom značne vyšší výkon.

K správnej funkcii HTTP potrebujeme však zložitejší model prihliadajúci aj na iné skutočnosti, ako tie, ktoré sme uvažovali doteraz. Potrebujeme vedieť niekoľko ďalších informácií, napríklad: celkovú štruktúru stránky, či je umiestnená na rovnakom serveri ako predchádzajúca požadovaná, a zväziť, že stránky sú rôznorodé a môžu sa navzájom obsahom a štruktúrou líšiť.

4.1 Súhrn existujúcich webových modelov

Pri vytváraní správneho modelu HTTP by sme mali vychádzať s adekvátnych meraní a rozboru skutočnej premávky a patrične prihliadať k jej požiadavkám. Poznáme tri základné metódy meraní pričom každá z nich vychádza z údajov čerpaných z iného miesta v sieti. Prvá priamo analyzuje záznamy z denníkov (nazývané tiež logy) na strane servera, druhá denníky na strane klienta a tretia metóda analyzuje údaje v silne frekventovanom mieste siete, kde dochádza ku všetkým druhom komunikácií.

1. Pozorovanie záznamov na strane servera

Opiera sa o poznatky získané z logu servera, ktorý si ukladá informácie o rozosielených súboroch, ktorými obsluhuje klientov. Z týchto informácií dokážeme vytvoriť jeho záťažový model. Nevýhoda tejto metódy spočíva v tom, že z nej nedokážeme odvodiť správanie samotného užívateľa, len dôsledky jeho správania na formu obsluhy.

2. Pozorovanie záznamov na strane klienta

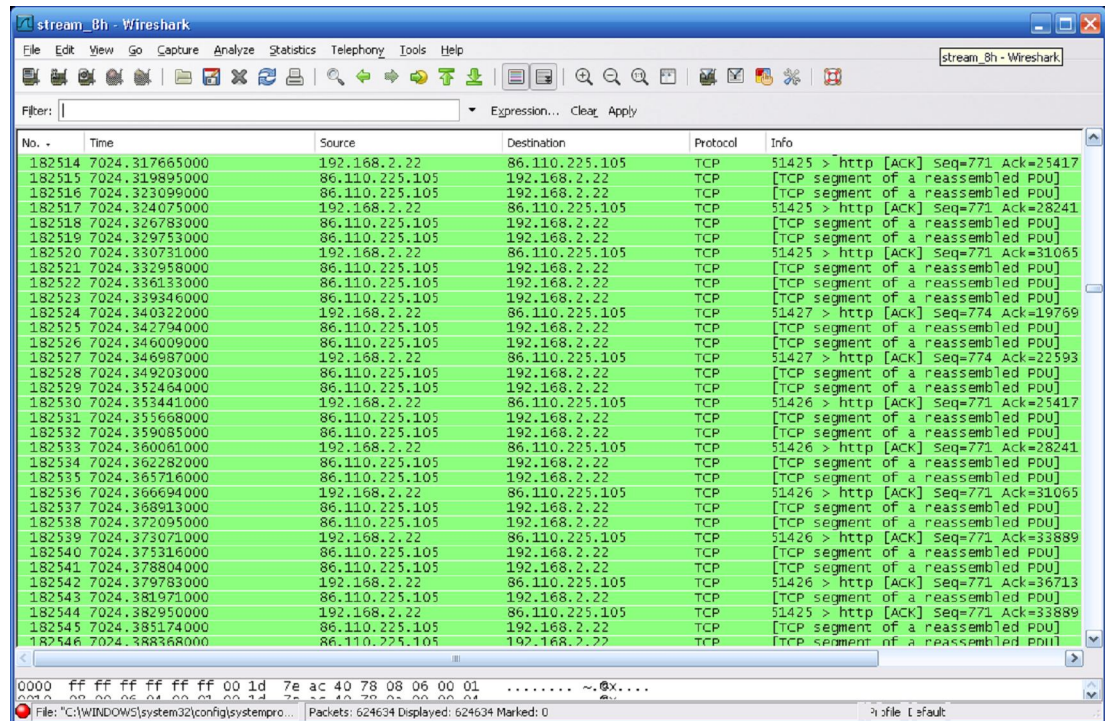
Opiera sa o poznatky získané z logu klienta. Pozorujú sa pri nej jednotlivé prístupy užívateľov k webu. Z týchto údajov dokážeme zostaviť odpovedajúci model správania sa WWW užívateľa. V súčasnej dobe sú však problémom zastarané poznatky zo štúdií týchto záznamov. Dnes sú zdrojové kódy najčastejšie používaných webových prehliadačov vo vlastníctve spoločností, ktoré ich vyvinuli. Človek na akademickej pôde sa k zdrojovým údajom tohto charakteru nemá možnosť dostať. Ďalšie ťažkosti vyplývajú s počtu ochotných užívateľov vybavených takto modifikovaným prehliadačom v súvislosti s ochranou osobných údajov. Aféru v tejto oblasti vyvolala v minulosti spoločnosť Google, ktorá používala vo svojich prehliadačoch Chrome monitoring užívateľmi prehliadaného obsahu. Na druhú stranu by mohli výsledky týchto rozsiahlych štúdií prispievať k vývoju prehliadačov a samotných protokolov.

3. Sledovanie samotného paketového prenosu

Meranie sa zakladá na sledovaní samotného paketového prenosu. Odchyťovaním paketov podsiete nesúcej HTTP prenos, typicky v sieti typu Ethernet, alebo ktorejkoľvek sieti, ktorá používa zdieľané médium prostredníctvom paketových analyzátorov (snifferov). Z odchytených dát sa pomocou programovej analýzy spracuje výsledný model premávky. Tento spôsob eliminuje nedostatky predošlých dvoch techník. Výsledkom sú detailnejšie údaje spracované priamo zo správania sa jednotlivých užívateľov. Treba však pripomenúť, že svojvoľné odchyťovanie dát za účelom ich analýzy je vo väčšine prípadov v rozpore s pravidlami prevádzkovateľa týchto sietí, nakoľko môže byť takáto činnosť chápaná ako porušenie súkromia a zneužívanie osobných údajov a môže mať trestno-právne dôsledky. Pre sledovanie siete na výskumné účely však môže byť realizovateľná. Príklady paketových snifferov a analyzátorov sú WireShark (Ethereal), TCPdump, Cain and Abel, Dsniff a ďalšie. Príklad odchytenia komunikácie medzi HTTP serverom a klientskeho PC pomocou programu WireShark je na Obrázku 1.

Prvými dvoma metódami sa už nebudeme ďalej v tejto práci zaoberať nakoľko sú zjavne neefektívne. Pripúšťame však, že z pohľadu komplexnosti je potrebné prihliadať aj k týmto údajom a vybrať si z nich potrebné užitočné informácie. Z modelu odvodeného z analýzy odchytených dát sa budeme zaoberať poznatkami uvedenými v literatúre [1], [5] a [12]. Je v nej opísaný model webového prenosu, ktorý sa nazýva model správania. Po pochopení modelu správania opierajúceho sa o poznatky užívateľov sa budeme snažiť navrhnúť svoj vlastný model, ktorý bude opisovať správanie sa užívateľov a ich požiadavky na sieťový prenos v HTTP a VoIP. Súbor odchytených

paketov programu Wireshark HTTP prenosu je uložený na CD médiu, ktoré je súčasťou odovzdania záverečnej práce.



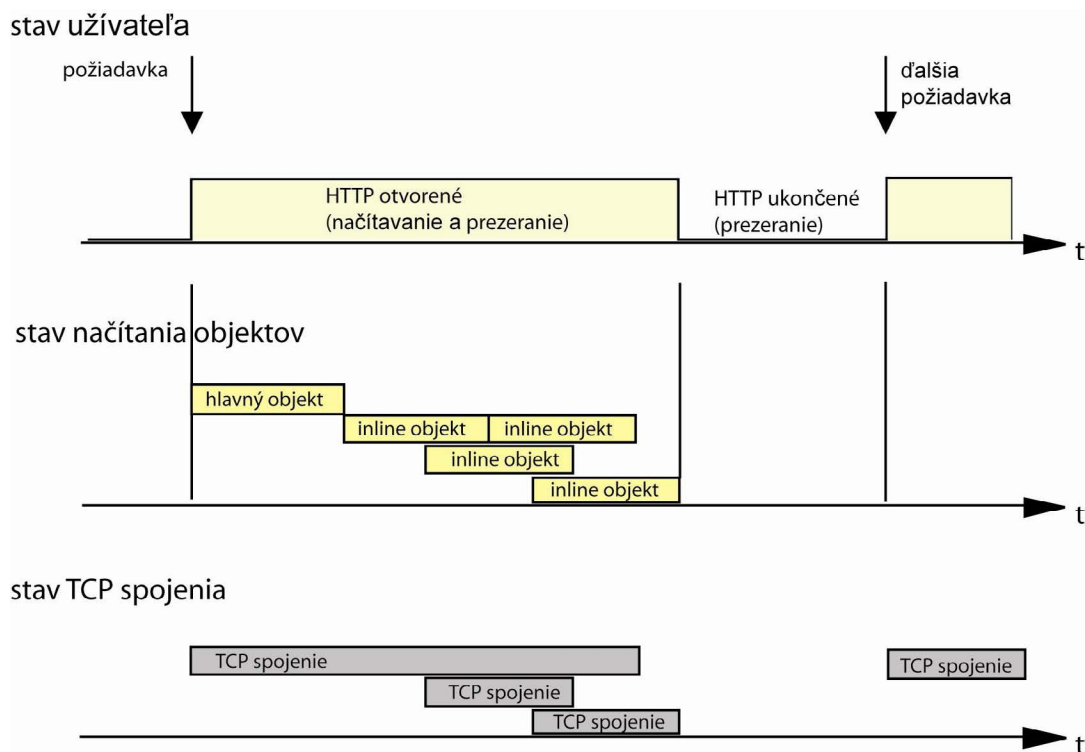
Obrázok 1. Príklad odchytenej komunikácie protokolu HTTP

4.2 Model správania

Model správania na Obrázku 2 podľa [1] graficky znázorňuje procesy prebiehajúce pri prehliadaní webu užívateľom. Jedná sa o súhrnný model HTTP premávky. Má hierarchickú povahu podobne ako modely zdrojovej premávky v sieťach ATM prezentované opakovaně v rôznych druhoch literatúr venovaných vysokorýchlostným sieťam. Vytváraný model bude značne podobný, založený na modeli správania sa užívateľa uvedený v [1]. Jeho aktivitu môžeme rozdeliť do niekoľkých premávkových vrstiev.

- 1. Vrstva HTTP požiadaviek:** Patrí sem proces prichádzajúcej požiadavky od užívateľa. Model jeho správania je založený na stavoch procesov ON/OFF, pričom stav ON prezentuje aktivitu po prijatí požiadavky až po odoslanie celého obsahu stránok, po ktorom nasleduje stav OFF. Trvanie týchto stavov je ovplyvnené rýchlosťou načítanie obsahu stránok a časom, v ktorom si ich užívateľ prezerá alebo premýšľa nad ich obsahom.

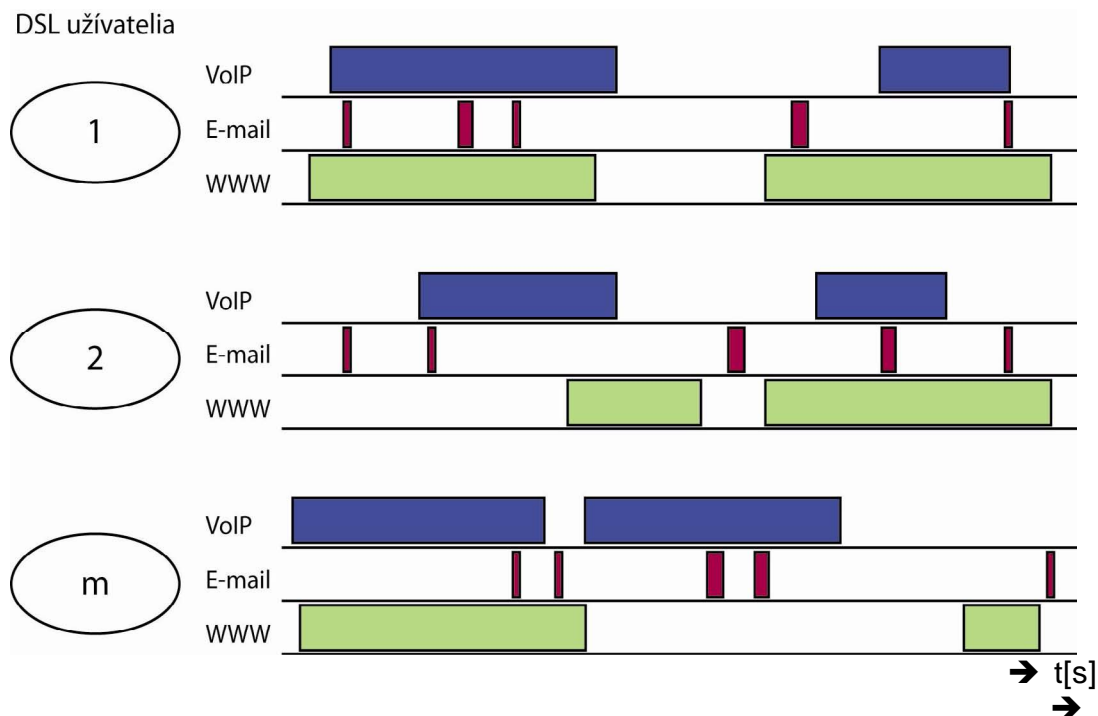
2. **Vrstva načítania objektov a ich prezerania:** (stav procesu ON) Predstavuje čas pre odosielanie všetkých objektov patriacich k jednej požiadavke. Užívateľ v tejto fáze čaká, kým sa objekty načítajú, môže však už priebežne prezerať doposiaľ načítaný obsah. Voliteľne môže proces prerušiť kliknutím na iný zobrazený odkaz a celý proces začína odznova.
3. **Vrstva prezerania:** (stav procesu OFF) Čas prezerania môžeme označiť ako neaktívny časový interval medzi dvoma požiadavkami. Po uplynutí doby prezerania sa generuje automaticky nová webová požiadavka.
4. **Vrstva načítania objektov:** Model štruktúry webového dokumentu považujeme za podobný ako v [5]. Bolo zistené, že veľkosť oboch objektov, hlavného aj inline objektu má logaritmicko-normálne rozdelenie. Zlepšenie modelu oproti modelu uvedeného v [5] spočíva v odvodení počtu inline objektov na jednu webovú stránku. V [1] bola autorom navyše analyzovaná veľkosť požiadavky *getrequest* s použitím logaritmicko-normálnej distribúcie.
5. **Aktivácia a deaktivácia TCP spojenia:** V stave ON sa aktivuje niekoľko TCP spojení, počas ktorých sú prenesené viaceré inline objekty. Záleží na prehliadači, či sú tieto spojenia otvárané paralelne alebo sériovo. Paralelné TCP spojenia pre jednotlivé inline objekty sú otvorené následne po otvorení prvého spojenia pre hlavný objekt stránky. Počet súčasne otvorených TCP spojení je závislý na verzii protokolu HTTP. Vo verzii 1.0 boli objekty sťahované postupne jeden za druhým a každý korešpondoval s jedným TCP spojením. V jeho modifikovanej verzii s viacerými TCP spojeniami prehliadač otvára viaceré spojenia a sťahuje tieto objekty paralelne. Počet týchto spojení je však limitovaný. Vo verzii 1.0 s *keep-alive* (udržiavaním spojenia) sa spojenia nezatvárajú okamžite, ale čaká sa na ďalšie možné požiadavky na prenos. V novej verzii 1.1 je povolené udržiavať konštantné spojenia a objekty posilať s ich pomocou.



Obrázok 2. Model správania sa užívateľa podľa [1]

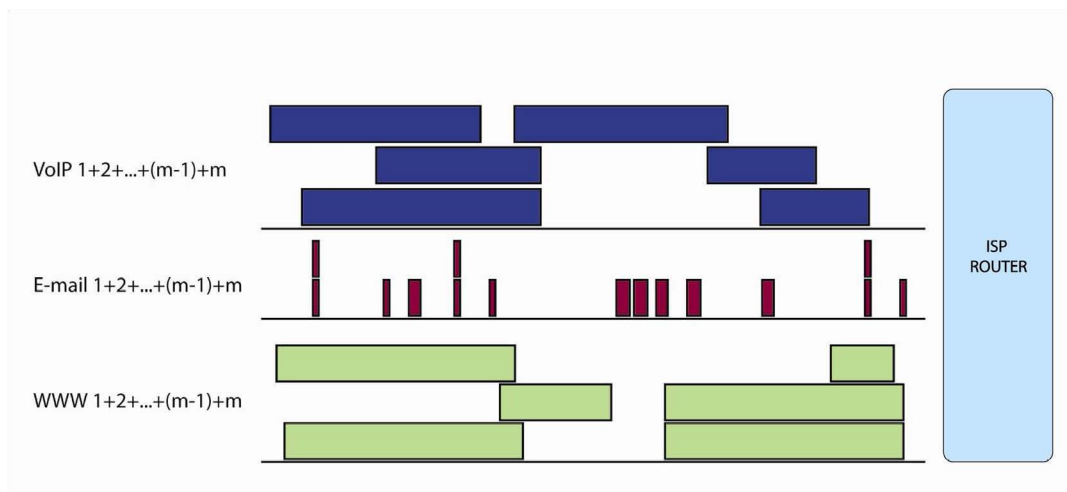
4.3 Navrhovaný model

V tejto časti je snaha navrhnuť model, ktorý by analyzoval priebeh sieťovej premávky z pohľadu technológie DSL a jej užívateľov. Cieľom nie je pochopiť DSL prenos ako taký, ale pomocou modelu sieťovej premávky odhadnúť o aký druh prenosu ide a prispôbiť sa k jeho požiadavkám. Do úvahy sa berú poznatky s kapitoly 3.2 aplikované na údaje získané pozorovaním počtu m DSL užívateľov s dátovým obmedzením. V dnešnej dobe je to stále najbežnejší predpokladaný druh pripojenia koncového domáceho užívateľa a firmy malého rozmeru do siete internet v SR (podľa portálu *dsl.sk*). Každý z týchto užívateľov generuje odlišné dátové toky závislé na konkrétnej aplikácii. Užívateľ môže v jednom časovom okamihu pracovať s viacerými aplikáciami, ktoré generujú svoje dátové toky paralelne (WWW, E-mail, VoIP, výmenné siete). Obráz takéhoto správania sa užívateľa je na Obrázku 3, kde je vidieť existenciu aktívnych fáz a typ dátových tokov pre m užívateľov.



Obrázok 3. Znáznornenie typu dátových tokov generovaných DSL užívateľom

Uvažované DSL linky sú zakončené v jednom spoločnom centrálnom uzle siete poskytovateľa internetového pripojenia, ktorý je realizovaný zariadením, v tomto prípade centrálnym smerovačom (router) ISP, ktorý pripája viacero bodov svojej hlavnej siete a teda aj DSL zákazníkov (pomocou prvku ISP router do siete internet. Situácia dátovej premávky v tomto uzle siete bude mať charakter graficky naznačený na Obrázku 4. Model výslednej základnej dátovej premávky je daný superpozíciou parciálnych prenosov a ich fáz pre všetkých užívateľov. Pre väčšiu prehľadnosť bol výber zúžený na tri dominantné typy prenosu ako webový prenos, služby elektronickej pošty a VoIP telefónie. V skutočnosti je týchto typov samozrejme viac.



Obrázok 4. Znáznorenie superpozície dátových tokov prichádzajúcich do centrálneho uzla siete poskytovateľa

4.4 Superpozícia dátových tokov viacerých užívateľov

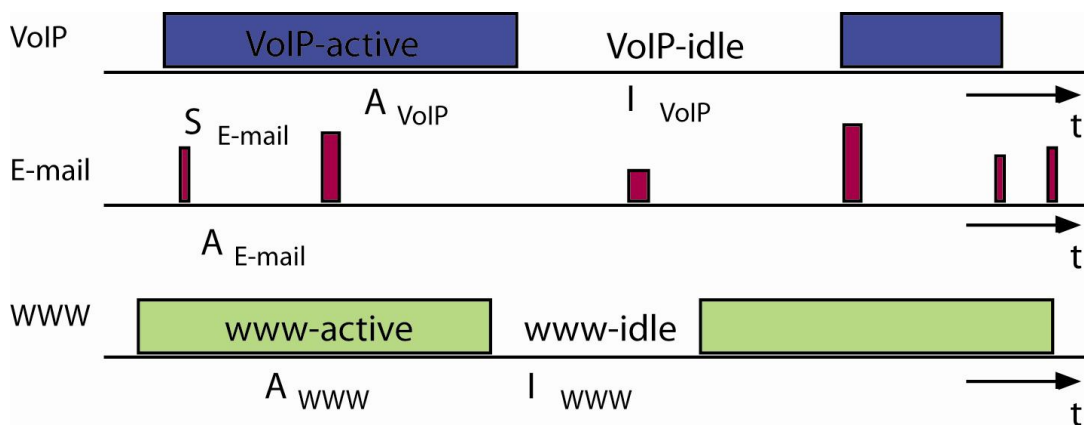
Model aplikačne orientovaného sieťového prenosu z pohľadu centrálneho uzla siete je na Obrázku 4. Výsledný dátový tok je daný superpozíciou všetkých dátových tokov jednotlivých užívateľov. Z telekomunikačného prenosového hľadiska môžeme výsledné dátové toky modelovať pomocou:

1. stavového procesu, ktorý zahŕňa zlučovanie m ON a OFF procesov užívateľov. Tieto stavy zodpovedajú výsledným stavom, ktoré prichádzajú od zdroja tak ako je zachytené na Obrázku 4. Pre výsledný pravdepodobnostný model premávky aplikácií WWW, HTTP a VoIP bude využitý stavový procesný model.
2. aproximujúceho stochastického obnovovacieho procesu (renewal process), ktorý je vytvorený zlúčením m procesov jednotlivých užívateľov. Jedná sa o aproximáciu konečného zdrojového procesu pomocou nekonečného zdrojového procesu.

4.5 Model prenosu jedného užívateľa

Prenos správania sa jedného DSL užívateľa je zobrazený na Obrázku 5. Podľa uvažovaného typu užívateľa budeme brať do úvahy nasledujúci opis jeho procesov :

1. Proces *ON/OFF* : Stav ON signalizuje aktívnu fázu (active, A_x , $x = \text{VoIP, E-mail, www}$), v ktorej užívateľ pracuje s aplikáciou, stav OFF je naopak perióda jeho nečinnosti (inactive, I_x , $x = \text{VoIP, E-mail, www}$). Trvania týchto stavov sú opísateľné náhodnými premennými a ich distribučnými funkciami. Príklad parametrizácie HTTP užívateľa sa pokúsime navrhnuť v ďalšom.
2. Stochastický obnovovací proces (renewal process) : Takýto proces je použiteľný ak stav ON je krátky v porovnaní s dĺžkou cyklu ON/OFF. Príkladom tohto typu procesu je E-mailová komunikácia popísaná ďalej.



Obrázok 5. Model správania sa jedného DSL užívateľa

4.6 Model HTTP

V kapitole 4.3 bolo uvedené, že výsledný tok dát pre webové prenosy v spoločnom prvku siete (napríklad už spomínaný router ISP) je daný superpozíciou jednotlivých dátových tokov všetkých aktívnych užívateľov, ak vezmeme do úvahy počet m aktívnych užívateľov prístupujúcich k HTTP pomocou jedného sieťového prvku súčasne. Obrázok 6. znázorňuje, že každý z nich bude opísaný stavom aktívnej A_{www} alebo nečinnnej I_{www} fázy:

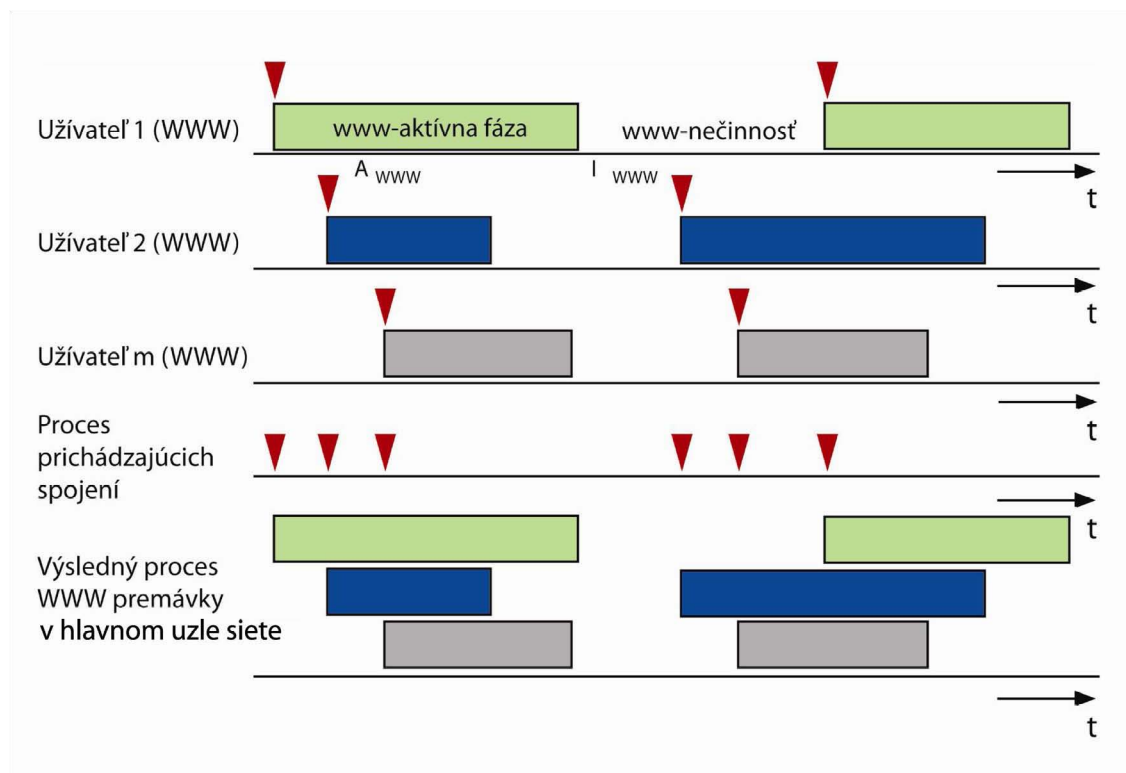
1. A_{www} prezentuje stav aktivity WWW. Táto fáza začína, keď užívateľ otvorí prehliadač, zadá prvú požiadavku a pokračuje ďalej v

prehliadaní bez dlhšieho prerušenia. Končí pri dlhšej fáze nečinnosti, keď užívateľ prestane HTTP používať. Prakticky sa tento stav končí vypršaním časovača.

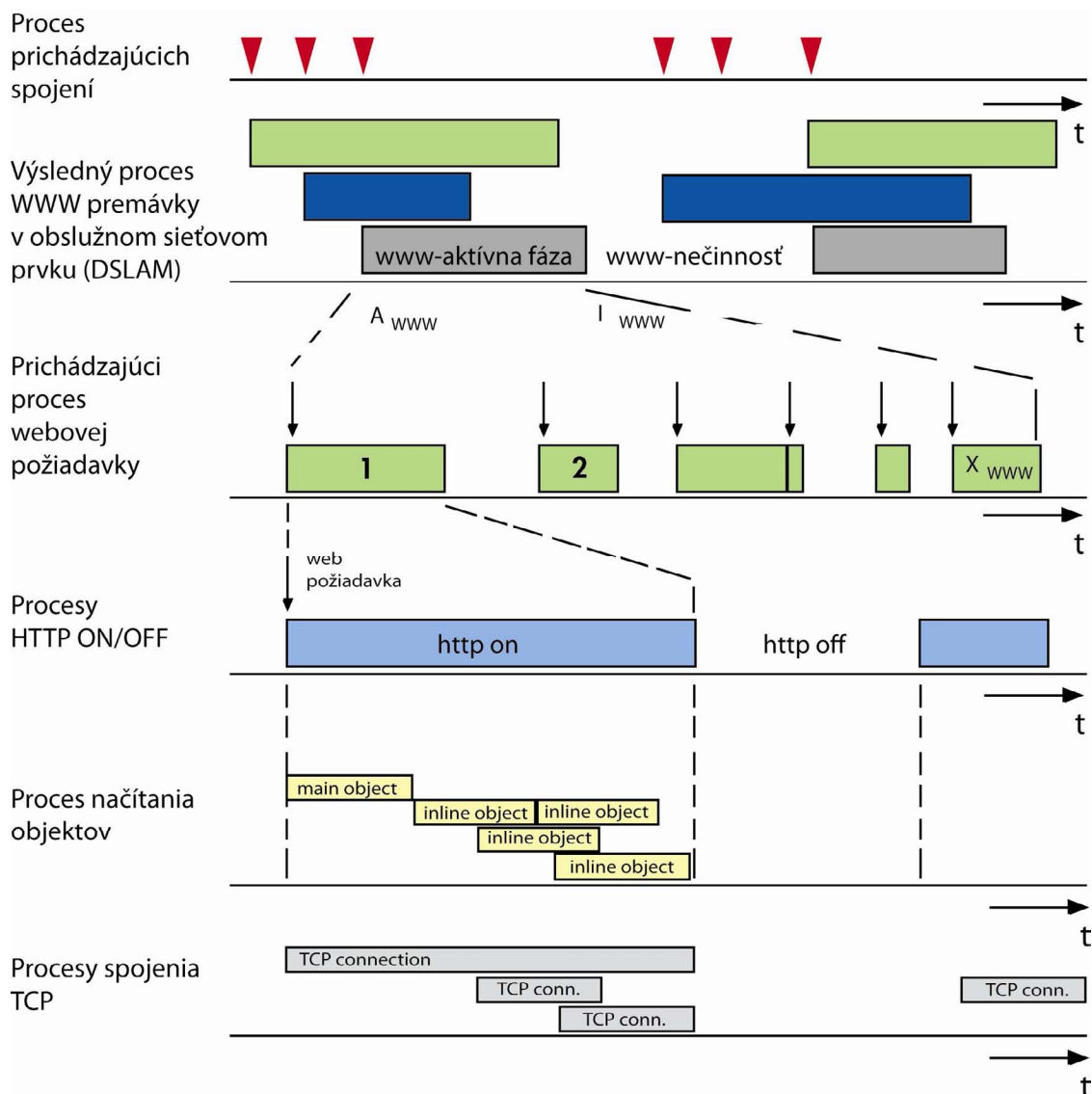
2. I_{www} prezentuje stav nečinnosti WWW.

Z pohľadu sieťového prvku na strane poskytovateľa rozlišujeme následné náhodné procesy súvisiace s premávkou WWW :

1. prichádzajúci proces spojení (alebo ich počiatkové body) s časmi medzi jednotlivými spojeniami T_{www}
2. kumulatívny stav procesov daný superpozíciou procesov ON/OFF jednotlivých užívateľov m . Výsledný stavový proces zodpovedá prichádzajúcim konečným zdrojovým procesom zobrazeným na Obrázku 6. Stav týchto procesov predstavuje počet súčasne sa nachádzajúcich užívateľov v aktívnej fáze WWW. Pre tento prípad je uvedený model správania WWW užívateľa počas jeho aktívnej fázy popísanej v ďalšom.



Obrázok 6. Výsledná HTTP premávka na hlavnom obslužnom zariadení (router ISP) [4]



Obrázok 7. Proces aktivity pre www [4]

Na Obrázku 7 je zachytený proces aktivity vo www, vychádzajúci z fáz činnosti HTTP ON a nečinnosti HTTP OFF uvedený v [1], [5], [12]. Fáza HTTP ON začína pri príchode web požiadavky. HTTP OFF fáza reprezentuje časový interval nečinnosti, ktorý nasleduje po tom, ako sú všetky požadované objekty prenesené. Oba sú závislé na časoch načítania a prezerania stránok, pričom čas prezerania môže byť teoreticky rovný aj nula vtedy, keď užívateľ odošle novú požiadavku ešte predtým ako sa načíta kompletný obsah stránok vrátane všetkých objektov. Počas fázy HTTP ON sa načítavajú všetky objekty, ktoré tvoria požadovanú stránku. Ako už bolo spomenuté skôr, rozlišujeme dva druhy týchto objektov: hlavný objekt (*main object*), pozostávajúci z hlavného HTML dokumentu a vedľajší objekt (*inline object*), ktorého odkaz je uložený v hlavnom HTML dokumente. V závislosti na použitom prehliadači

môže byť pre potreby načítania týchto objektov súčasne aktivovaných niekoľko TCP spojení. Aktivácia a deaktivácia nového TCP spojenia môže prebiehať buď sériovo alebo paralelne. Paralelné TCP spojenie môže byť otvorené pre inline objekt následne po nadviazaní hlavného spojenia pre main objekt.

Výsledný model pozostáva z dvoch častí. Prvá časť pozostáva z parametrov opisujúcich časy medzi dvomi reláciami (session) a počtom webových požiadaviek na jednu reláciu. Druhá časť opisuje zloženie webovej požiadavky určenej veľkosťou hlavného objektu a niekoľkými ďalšími inline objektmi, ich veľkosťou, veľkosťou požiadavky *getrequest* a dĺžkou prezerania jednej stránky. Model teda pozostáva z nasledujúcich pravdepodobnostných parametrov :

I_{WWW}	čas medzi reláciami WWW
X_{WWW} :	počet požiadaviek na jednu reláciu
V_{WWW} :	čas prezerania stránky
M_{WWW} :	veľkosť hlavného objektu
N_{WWW} :	počet inline objektov
O_{WWW} :	veľkosť inline objektov
R_{WWW} :	veľkosť <i>getrequest</i>

Rozdelenia pravdepodobnosti, stredné hodnoty a smerodajné jednotlivých parametrov boli vyhodnotené na základe meraní a zistení uvedených v [1], [20] sú v Tabuľke 1 (rozdelenia pravdepodobnosti, ich stredné hodnoty a smerodajné odchýlky $E()$ a $\sigma()$ sú uvedené v kapitole 8.1 Vybrané pojmy z teórie pravdepodobnosti a matematickej štatistiky v závere práce) :

Parameter	Rozdelenie	Stredná hodnota	Smerodajná odchýlka
I_{WWW}	Exponenciálne	-	-
X_{WWW}	Lognormálne	25 stránok	100 stránok
V_{WWW}	Weibullovo	39.5s	92.6s
M_{WWW}	Lognormálne	10 kB	25 kB
N_{WWW}	Gama	5.55	11.4
O_{WWW}	Lognormálne	7.7 kB	126 kB
R_{WWW}	Lognormálne	360 B	106 B

Tabuľka 1. Hodnoty parametrov pre model HTTP

Doba I_{WWW} medzi reláciami WWW sa dá určiť iba z dostatočne dlhého merania. Možno tu však predpokladať štatistickú nezávislosť a časy medzi jednotlivými reláciami je možné modelovať exponenciálnym rozdelením a prichádzajúce relácie považovať za Poissonov proces. Parametre uvedené v Tabuľke 1 zodpovedajú priemernej veľkosti webovej požiadavky približne 50 kB.

5. Model prenosu hlasu a videa VoIP

V tejto dobe sa prenos hlasu cez IP (VoIP, Voice over Internet Protocol) dostáva čoraz viac do popredia a jeho popularita rastie najmä vďaka priaznivej cene oproti klasickej telefónii a rozvoju infraštruktúry sietí s prepínaním paketov a multimediálnych aplikácií.

VoIP si získala svojich používateľov práve na základe skutočnosti, že im dovolila komunikovať s využitím IP protokolu a internetu namiesto telefónnych sietí PSTN. V posledných rokoch sa situácia stala ešte viac pre VoIP priaznivá, začali vznikať nové „open source“ platformy aj v spojení s výmennými sieťami (napríklad Asterisk), výrobcovia zariadení rozšírili svoje ponuky o telefónne adaptéry priamo podporujúce technológiu VoIP a taktiež poskytovatelia pripojení do siete Internet (ISP) pre bežných užívateľov rozšírili balíky služieb umožňujúce práve komunikáciu cez VoIP.

VoIP technológie sa často masovo využívajú na firemných úrovniach, ako efektívne riešenia telefónnej komunikácie z hľadiska ceny. Aplikácie VoIP sú vo vlastníctvach jednotlivých spoločností, ktorých zamestnanci môžu s výhodou komunikovať medzi sebou či už s počítača na počítač prípadne z počítača na telefón s využitím špeciálnej aplikácie a doplnkovej slúchadlovej sady s mikrofónom. Výnimkou nie je ani bezdrôtový WiFi telefón.

Samotné siete s prepínaním paketov s využitím IP protokolu nemôžu zo svojej podstaty vyhovieť požiadavkám na oneskorenie pri hlasovej komunikácii, preto VoIP aplikácie požadujú riadenie siete s podporou QoS. Projekty v súčasnosti sa sústreďujú na budovanie infraštruktúry týchto sietí, no veľmi málo je takých, ktoré skúmajú dopad VoIP na sieť ako takú a modelovanie časti prenosu venovanému priamo VoIP.

Programov a utilít pre priame monitorovanie premávky VoIP v súčasnosti nie je veľa. Najčastejšie sú integrované do paketových skenerov ako je napríklad WireShark a používajú sa skôr pre vyhľadávacie účely v špecifických situáciách (nekompatibilita kodekov, strata paketov a pod.). Tieto sú väčšinou nevhodné na stále monitorovanie VoIP premávky. Ďalšie ako Vomit [14] alebo RTP Tools [10] sú vhodné viac pre zachytenie hlasovej komunikácie ako takej alebo iné na zabezpečenie siete pre odhaľovanie útokov škodlivého softvéru. Príklad systému, ktorým bola monitorovaná VoIP sieť veľkej spoločnosti uvedený v [2] pre účely štatistického modelovania VoIP komunikácie, spracovával „Call Detail Records“, ktoré sú súčasťou systému Cisco Call Manager. Výsledky dosiahnuté v tejto analýze sú uvedené v kapitole 5.4.

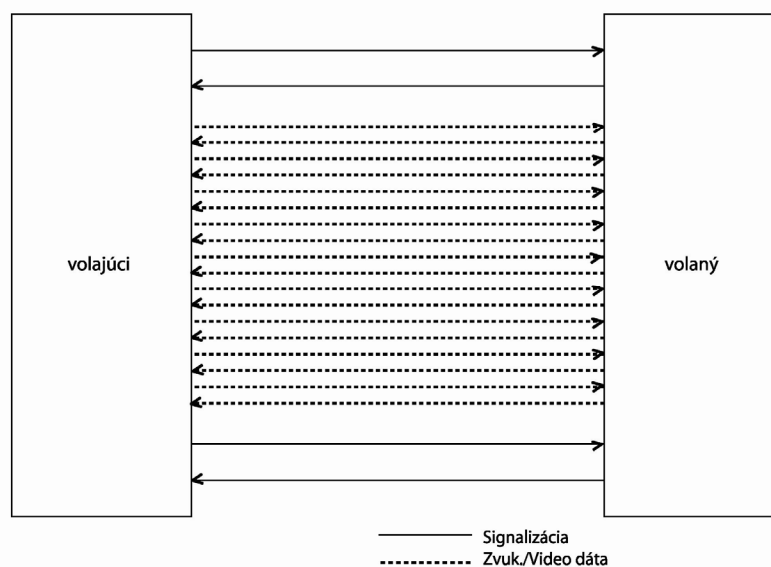
5.1 Základný model komunikácie dvoch užívateľov

V súčasnosti poznáme tri hlavné skupiny VoIP protokolov :

- štandardné protokoly ako SIP/H.323/RTP
- proprietárne, dobre zdokumentované protokoly (napr. Cisco 11 [11])
- proprietárne a nezdokumentované (napr. Skype)

Používanie štandardných protokolov však v praxi ešte neznamená, že je možné sledovať priebeh celej komunikácie ako napríklad RTP, ktoré pri prenose môže mať zvukové a video dáta zakódované súkromnými kodekmi.

Pre modelovanie prenosu sa v práci vychádza z podstaty prenosu ako takého, ktorý je založený na protokoloch bez spájania, akým je UDP.



Obrázok 8. Model komunikácie 2 užívateľov prostredníctvom VoIP [15]

5.2 VoIP protokol - architektúra

Obrázok 8 znázorňuje základ VoIP. Každá komunikácia sa skladá z niekoľkých základných krokov.

Ak chce volajúci komunikovať s volaným, musí nadviazať spojenie buď priamo s volaným alebo s gateway/PBX (záleží, aký protokol sa bude používať

a na nastavení lokálnej siete) s využitím signálneho protokolu. Pri tomto kroku prebieha:

- overovanie (kredit, schopnosť telefonovať, atď.)
- dohadovanie typu hovoru, či sa jedná o hlas, prípadne kombináciu hlasu/video
- zvolenie spoločného kodeku, (napríklad H.264)
- dohodnutie sa na portoch, ktoré sa budú používať

Nasledujúca Tabuľka 2 zobrazuje najpoužívanéjšie signalizačné a transportné protokoly využívané pre VoIP.

Signalizačné protokoly VoIP	Transportné protokoly VoIP
○ SIP (Session Initiation Protocol)	○ RTP (Real-time Transport Protocol)
○ Cisco 11	○ RTCP XS (RTP Control Extended Reports)
○ H.323	

Tabuľka 2. Signalizačné a transportné protokoly VoIP

Z hľadiska pozorovania pri modelovaní sieťového modelu:

- Signalizačný protokol pozostáva z veľkého počtu dôležitých informácií ako napríklad totožnosti jednotlivých účastníkov, typu hovoru (hlasový alebo video hovor), použitých kodekov, trvaní a informáciách o jednotlivých reláciách, ktoré zvyčajne neprebiehajú na tých istých portoch. Bez správneho dekódovania signalizačného protokolu, nie je možné definovať ktoré porty RTP sa využijú.
- Transportný protokol pre zvuk a video sa používa na získanie informácií ako chybovosť, strata paketov, oneskorenie a z nich sa odhadujú hodnoty určujúce kvalitu hovoru.
- Protokoly ako RTCP XS RTP (RTP Control Extended Reports) sa používajú na prenos informácií o RTP tokoch vrátane informácií ako strata paketu, oneskorenie, metriky prenosu hovoru. Tieto informácie zvyčajne nebývajú dostupné pre VoIP hovory, pretože sú často generované jeho príslušenstvom.

Pre ostatné uzavreté a proprietárne protokoly ako Skype nie je možné dekódovať hovor z viacerých dôvodov. Napríklad Skype je založený na protokole eDonkey P2P a v mnohých prípadoch prenosu je problém zistiť, či vôbec ide o hlasový alebo video hovor.

5.3 Sieťová premávka

V závislosti od použitého protokolu ju delíme do dvoch skupín: proprietárnu a štandardnú VoIP premávku. Proprietárna sa v tejto práci nebude ďalej uvažovať a preberaná bude štandardná VoIP premávka, ktorá sa ako už bolo uvedené v predchádzajúcej kapitole delí na premávku signálnych protokolov a premávku hlasových protokolov, kde sa používa v prevažnej väčšine protokol RTP, pôvodne vyvinutý za účelom flexibility a prenosu rôznych typov dát (nielen hlasu a videa) s kódovaním do viacerých formátov.

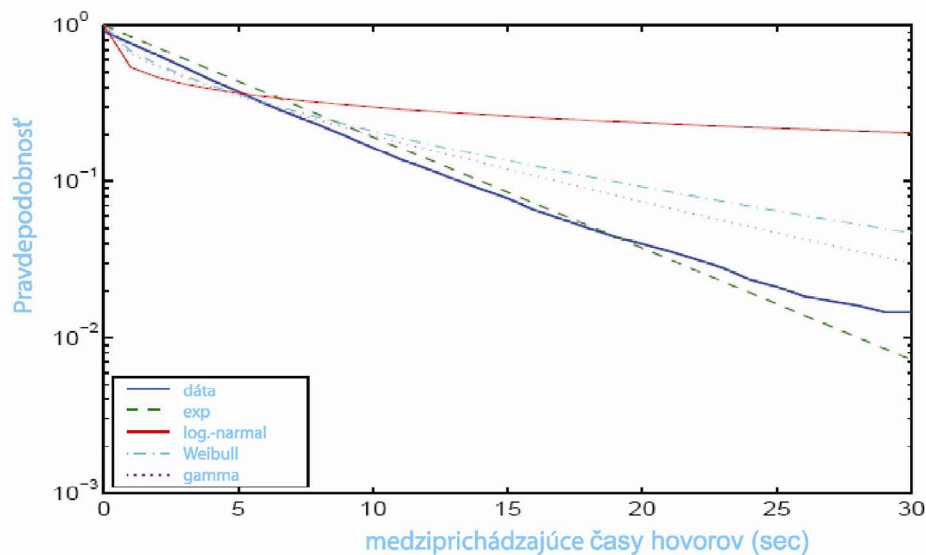
5.4 Model sieťovej premávky VoIP

Sieťová premávka VoIP na úrovni volania pozostáva z dvoch príbuzných procesov. Jedným je proces prichádzajúcich hovorov charakterizovaný dobou medzi dvoma po sebe idúcimi hovormi. Druhým je proces držania hovorov určený dobou jeho trvania. Oba procesy je možné samostatne analyzovať a navrhnúť osobitý model.

Hovor samotný, prebiehajúci počas jeho držania, na ktorý je aplikovaný kodek, je možné skúmať taktiež samostatne. Pozostáva z rečovej a kludovej fázy, ktoré sú udávané v literatúre približne 350 ms pre reč a 650 ms pre klud, čo zodpovedá konštantnej rýchlosti jeden paket za 16 ms vo fáze reči (ON).

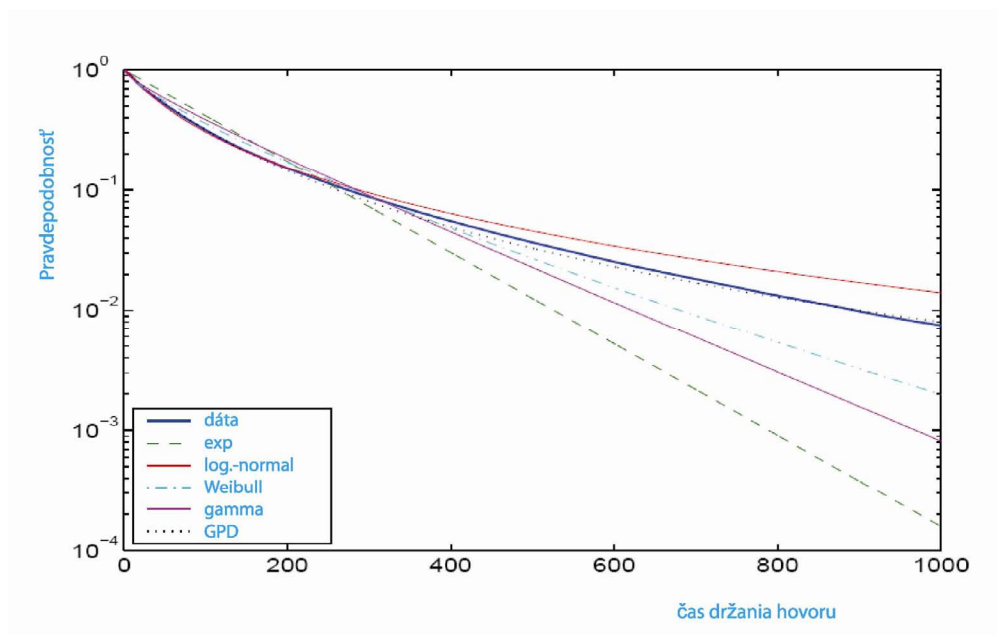
Autori v [2] popisujú sledovanie komunikácie VoIP v rozsahu periódy 240 pracovných dní, počas ktorej vyhodnocovali proces prichádzajúceho hovoru a proces držania hovoru v rušnej sieti veľkej spoločnosti (800 VoIP telefónov), kde bola premávka počas pracovnej doby takmer neustála. Informácie pre prichádzajúce hovory a držania hovorov poskytla databáza záznamov „Call Detail Records“ systému Cisco Call Manager Release 3.3. Analýza bola vykonaná v prostredí Matlab s využitím funkcií Wafo toolboxu. Do spracovania pre prichádzajúce hovory boli zahrnuté reprezentatívne rušné časti pracovných 211 dní, keď premávku možno považovať za stacionárnu. Držanie hovorov sa považovalo za nezávislé počas celých sledovaných 240 dní.

Štatistické spracovanie nameraných údajov vyvedené do semilogaritmických grafov poskytlo rôzne stochastické správanie pre oba procesy. Obrázku 9 ukazuje, že proces prichádzajúcich hovorov sa najviac približuje exponenciálnej rozdeleniu pravdepodobnosti s parametrom $\lambda = 0,164$. Autokorelačná funkcia vytvorená z nameraných dát ukázala, že prichádzajúce hovory môžeme považovať za nezávislé a teda Poissonov model dobre vystihuje proces prichádzajúcich hovorov.



Obrázok 9. Pravdepodobnosti časov prichádzajúcich hovorov [2]

Prekvapivý výsledok bol dosiahnutý v prípade vyhodnotenia držania hovorov. Klasické modely pre VoIP zvyčajne predpokladajú exponenciálne rozdelenie pre tento proces. Výsledok znázornený na Obrázku 10 však ukazuje, že koniec komplementárnej distribučnej funkcie časov držania hovorov klesá oveľa pomalšie ako pri exponenciálnej distribučnej funkcii, čo je zvyčajne charakteristické pre rozdelenia s tzv. ťažkými chvostmi. Namerané údaje boli vyrovnávané lognormálnym, Weibulovým, gamma a Paretovým rozdelením. Najlepšiu zhodu s údajmi vykázalo všeobecné Paretoovo rozdelenie (Definícia 8.14) s parametrami $GPa(-0.39, 69.33, 0.0)$ s označením GPD na Obrázku 10.

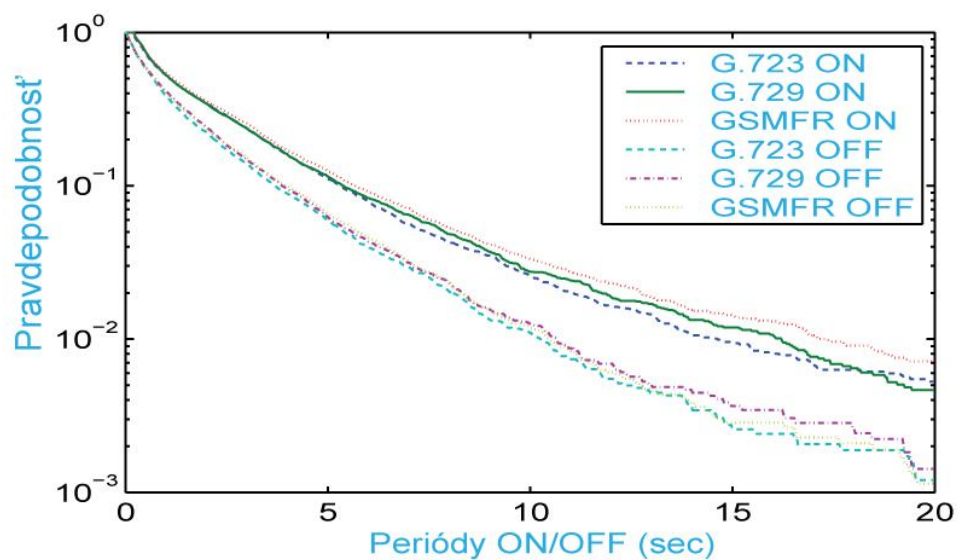


Obrázok 10 Pravdepodobnosti časov držania hovorov [2]

Uvedená analýza ukazuje, že proces prichádzajúcich hovorov je možné modelovať Poissonovým rozdelením pravdepodobnosti počtu hovorov, prípadne exponenciálnym pre doby medzi prichádzajúcimi hovormi. Navrhovaný model pre proces doby držania hovorov vyhovuje všeobecnému Paretovému rozdeleniu pravdepodobnosti.

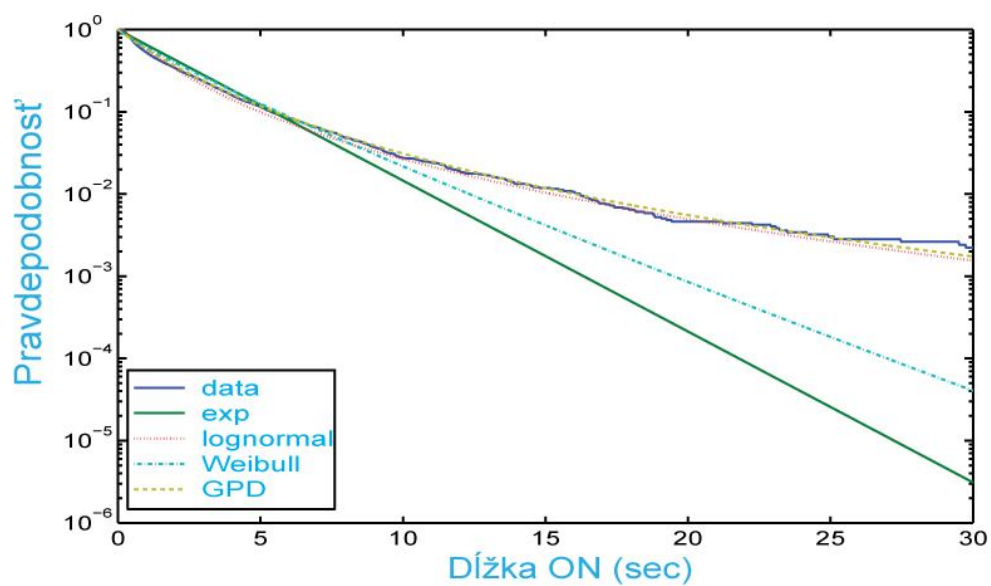
Ako bolo uvedené na začiatku tejto kapitoly VoIP je možné skúmať na úrovni samotného prenosu hlasových paketov (voice stream), čo bolo taktiež publikované v [2]. Skúmali sa tri kodeky (G.723, G.729B a GSMFR) s detekciou hlasovej aktivity (VAD) a zo zistených veľkostí periód ON/OFF boli zostavené komplementárne distribučné funkcie uvedené na Obrázku 11.

Z tesnosti zobrazených kriviek vyplýva, že kodeky mali veľmi malý vplyv na charakteristiky rozdelenia periód ON/OFF, čo možno zovšeobecniť na všetky kodeky rovnakých parametrov ako skúmané. Naviac Obrázok 11 poukazuje na skutočnosť, že grafické zobrazenie periód ON/OFF nemá lineárny priebeh v semilogaritmickej mierke, čo by zodpovedalo exponenciálnemu rozdeleniu pravdepodobnosti, ktoré navrhujú klasické modely pre VoIP. Skúmaná zhoda s rôznymi rozdeleniami pravdepodobnosti pre kodek G.729 ukázala najlepšiu zhodu pre všeobecné Paretovo rozdelenie (Definícia 8.14) s parametrami $GPa(-0.28, 1.7, 0.0)$ pre ON a s parametrami $GPa(-0.35, 1.02, 0.0)$ pre OFF periódu.

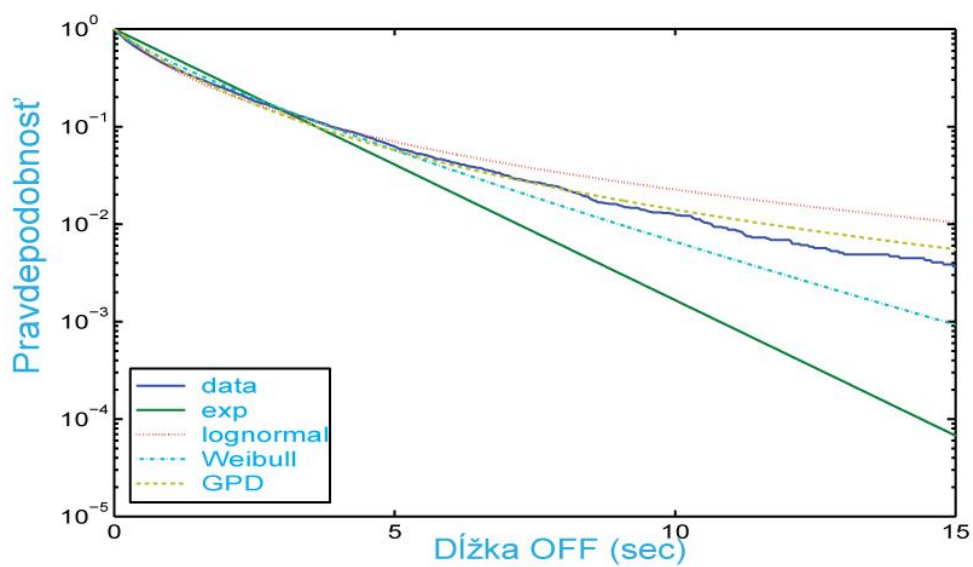


Obrázok 11. Pravdepodobnosti periód ON/OFF [2]

Zistenie, že na ON/OFF periód y zdroja hlasu vo VoIP nie je vhodné aplikovať agregačné modely založené na exponenciálnom rozdelení, ale presnejšie sú modely vychádzajúce z rozdelení s tzv. ťažkými chvostmi ako v tomto prípade Pareto ako je ilustrované na Obrázku 12 a 13. Z tohto poznania je v [2] urobený záver, že ON/OFF agregované zdroje so zisteným pravdepodobnostným y vlastnosťami majú tzv. vnútornú podobnosť (self-similarity) a fraktálne charakteristiky rozpracované v teórii chaosu. V dôsledku adaptačnej vlastnosti protokolu TCP je možný prenos fraktálnych vlastností agregovaných dátových tokov VoIP aj na iné dátové toky v sieti. Tento predbežný záver však predmetom ďalšieho skúmania.



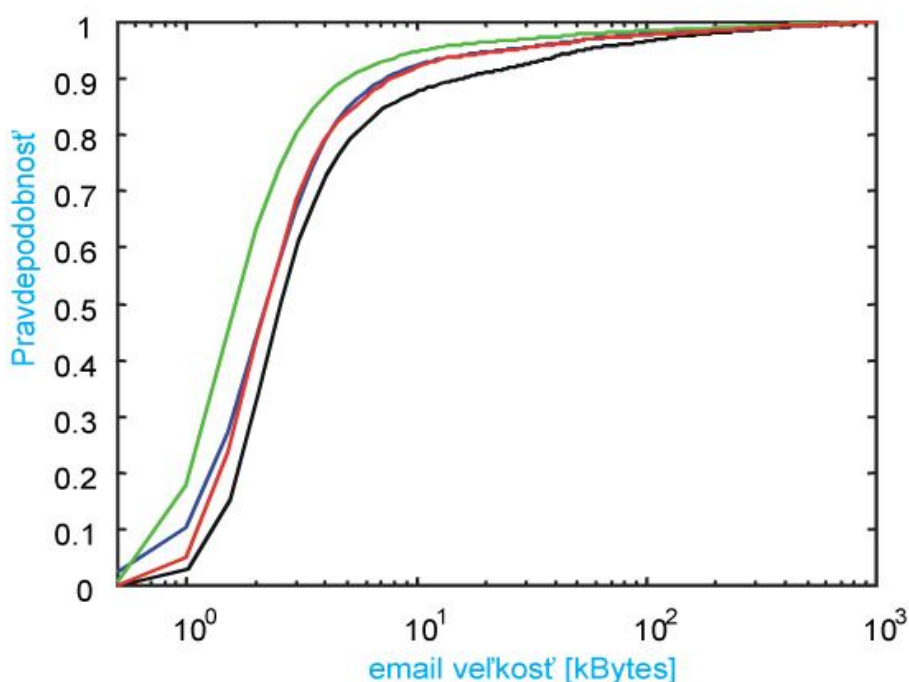
Obrázok 12. Pravdepodobnosti periódy ON pre kodek G.729 [2]



Obrázok 13. Pravdepodobnosti periódy OFF pre kodek G.729 [2]

5.5 Model elektronickej pošty

Elektronická pošta napriek tomu, že je staršia v porovnaní webovými aplikáciami a internetovou telefóniou stále zostáva veľmi často využívaným spôsobom komunikácie najmä v komerčnom sektore. Poskytuje veľmi jednoduchý a efektívny spôsob výmeny informácií. Najčastejšie sledovanou charakteristikou, ktorá sa štatisticky vyhodnocovala v citovanej literatúre je veľkosť E-mailu. Štatistické súbory údajov pochádzajúci z E-mailovej komunikácie troch užívateľov, každý obsahujúci viac ako 6000 E-mailov boli spracované v [4] a ich distribučné funkcie sú uvedené na Obrázku 14 a číselné údaje sú v Tabuľke 3.



Obrázok 14. Distribučné funkcie veľkostí E-mailov [4].

užívateľ	počet E-mailov	stredná hodnota [kB]	štandardná odchýlka [kB]
1	12110	17,6	174,8
2	7026	28,2	234,9
3	6787	30,1	209,6
> 3	34670	22,7	200,3

Tabuľka 3. Hodnoty zistených veľkostí E-mailov

Pre veľké E-maily Cauchyho rozdelenie pravdepodobnosti $C(0,8;1,0)$, (Definícia 8.17) dobre vystihuje meraním zistené údaje. Pre malé E-maily sú takto vyrovnané pravdepodobnosti nadhodnotené. Spôsobené je to tým, že E-mail má minimálnu dĺžku určený hlavičkou správy. Z porovnania údajov vyvedených v Obrázku 14 je zrejмый totožný tvar kriviek, prislúchajúce údaje sa odlišujú strednou hodnotou a štandardnou odchýlkou, čo je spôsobené rozdielnou aktivitou sledovaných užívateľov.

6. Softwarová simulácia modelu prenosu

Hlavným cieľom samostatnej softwarovej analýzy je zostavenie a porovnanie referenčného sieťového modelu v prostredí programu OPNET IT Guru Academic Edition a reálne pracujúcej firemnej siete malého rozmeru. Program OPNET IT Guru Academic Edition použijem pre zostavenie referenčného modelu počítačovej siete v ktorej prebiehajú prenosy HTTP a E-mail na základe uvedených teoretických poznatkov o pravdepodobnosti ich výskytu v prenose. Tento model porovnam so štatistickými údajmi zistenými programom WireShark a analyzátorom CACE Pilot v skúmanej reálne pracujúcej počítačovej sieti. Ako reálnu sieť som využil LAN spoločnosti, v ktorej som zamestnaný a ktorej som predpokladal, že poskytne premávku aplikácií HTTP a E-mail v dostatočnej miere na vytvorenie požadovanej analýzy.

6.1 Zostava reálnej počítačovej siete

Skúmaná sieť je malého rozsahu pozostávajúcej z približne 10 koncových zariadení zapojených v sieti LAN technológiou Ethernet s priepustnosťou 100 Mbit.s^{-1} s metalickými káblami STP. Pre potreby bezdrôtového spojenia je k dispozícii bezdrôtová LAN. Sieť má topológiu v tvare 2 x hviezda. Je umiestnená v trojpodlažnej budove a slúži potrebám spoločnosti pôsobiacej v oblasti reklamy, Obrázok 15 v prílohe. Základným sieťovým prvkom je tu 16 portový prepínač ASUS GX1016D. Pre podporu bezdrôtovej technológie 802.11.b/g je k dispozícii prístupový bod 3Com Office Connect Wireless g Access Point. Sieť je pripojená do siete internet pomocou ADSL smerovača Micronet SP916GK Wireless Broadband Router. Pre potreby analýzy toku dát bol pridaný medzi prepínač a smerovač 4 portový HUB 3Com Officeconnect a PC s programom WireShark pre odchytyvanie paketov a vytváranie štatistík, Obrázok 16 v prílohe. V sieti prebiehajú najviac prenosy aplikačných protokolov HTTP, SMTP a POP3, Obrázok 17 (CACE). Vyplyva to z charakteru zamerania činnosti spoločnosti, keď jednotliví pracovníci firmy pri komunikácii s klientmi najviac vytvárajú cenové ponuky, prehliadajú pri tom elektronické on-line katalógy dodávateľov a posielajú vo veľkej miere elektronickú poštu. Prenosy VoIP sa vo firemnej sieti nevyskytujú

nakoľko je v budove spoločnosti nainštalovaný telefónny systém DECT. Uvedené skutočnosti budú zohľadnené pri vytváraní referenčného modelu pri neskoršej softwarovej analýze.

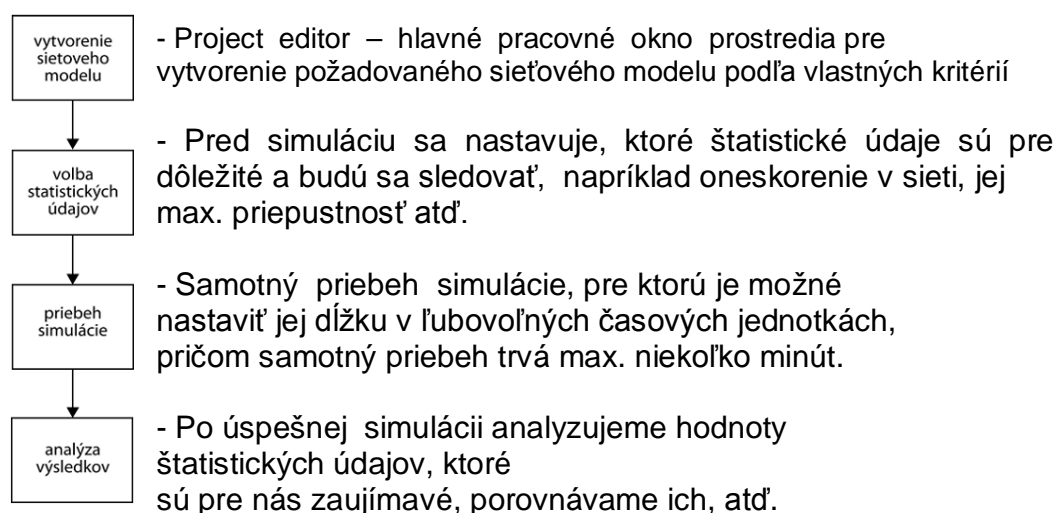
6.2 OPNET IT Guru Academic Edition 9.1

Program bol stiahnutý zo stránok výrobcu a nainštalovaný na PC v skúšobnej verzii, ktorá obsahuje oproti komerčnej verzii niekoľko obmedzení vo funkčnosti, v použití a v dostupnej dokumentácii. Je primárne určená na výučbu simulácie a analýzy sieťových modelov a ich topológie za pomoci vyhodnocovania štatistických údajov prebiehajúcich prenosov definovaných aplikácií. Po nainštalovaní je k dispozícii základné menu programu, Obrázok 18.



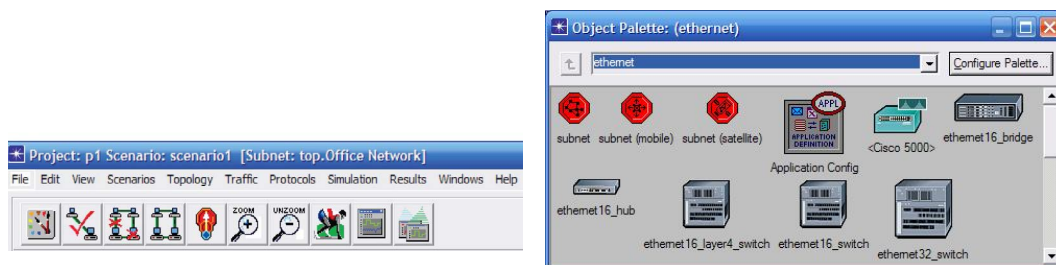
Obrázok 18. Menu programu OPNET IT Guru

Práca v programe IT Guru postupuje podľa nasledovnej schémy:



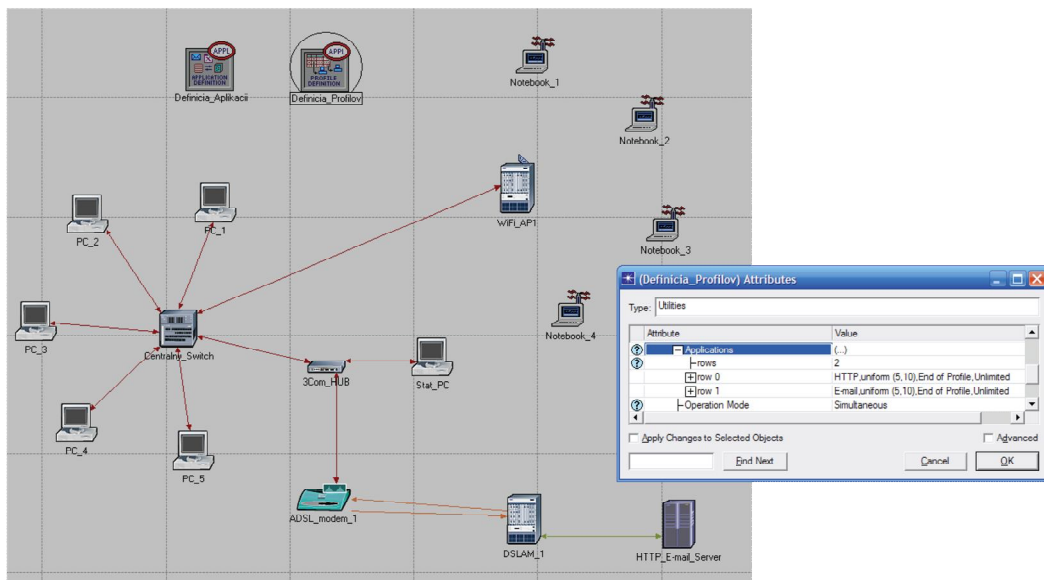
6.3 Postup vytvorenia referenčného modelu siete

V menu programu zvolíme nový projekt a jeho prostredie, ktoré najviac zodpovedá našim požiadavkám. Každý projekt môže obsahovať početne veľa prostredí na celosvetovej úrovni. Ďalej si volíme základnú topológiu siete, rozsah, mapu, veľkosť a plochu siete. Využívam štandardne plochu 100 x 100 metrov, ktorá simuluje rozlohu firemnej siete vrátane jej podlaží. Nasleduje ešte pridanie palety objektov do modelu siete (voľba knižníc objektov LAN, WLAN, xDSL). V ďalšom kroku sa vkladajú jednotlivé objekty podľa Obrázku 19 z palety podľa Obrázku 20 (6x PC, 4x notebook, prepínač, smerovač, prístupový bod) podľa topológie reálnej siete z Obrázku 15 uvedeného v prílohe. Koncové zariadenia sú medzi sebou spojené linkami 100baseT.



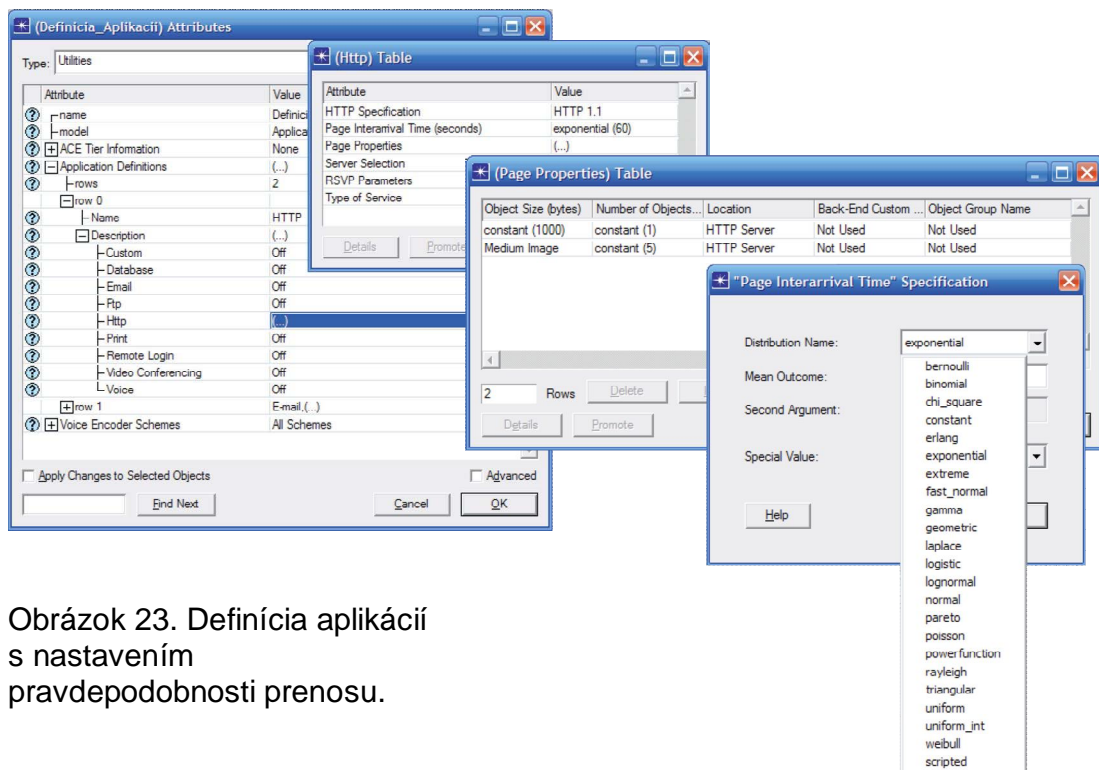
Obrázok 19. Menu Project manager. Obrázok 20. Paleta prvkov knižnice ethernet.

Pri zostavovaní modelu v obláčiku internet vytvoríme zjednodušené nereálne prostredie, ktoré bude schopné obsluhovať klientov aplikácií HTTP a E-mail službami a simulovať tak premávku v sieti. Boli vybrané prvky DSLAM s linkami ADSL upstream / downstream a server poskytujúci služby pre aplikácie HTTP a E-mail. Táto časť siete nie je reálna a je vytvorená pre potreby simulácie. Po takto zostavenom modeli siete si overíme správne zapojenie a funkčnosť liniek. Program po overení vyhodnocuje linky ako správne zapojené. Model siete treba doplniť o dôležité bloky definícií aplikácií, definíciu profilov pre simuláciu premávky v sieti. V simulácii referenčného modelu budú prebiehať prenosi aplikácií HTTP a E-mail, tak ako v analyzovanej reálnej sieti. Dôvodom je možné porovnanie ich správania po realizácii softwarovej analýzy a simulácie oboch sietí. Vytvorený projekt, Obrázok 19, je pripravený pre nastavenie aplikácií a aplikačných profilov simulujúcich premávku v sieti. Vhodným zariadeniam sa dá priradiť klient aj server aplikácie. Pre potreby simulácie je možné nastaviť odlišné správanie aplikácií a vytvoriť tzv. scenáre. Scenár obsahuje rovnakú alebo podobnú konfiguráciu tej istej siete s odlišnými parametrami správania aplikácií (Obrázok 20 v prílohe). Typicky sa pre každý scenár vytvorí simulácia a dosiahnuté výsledky sa v konečnej fáze vyhodnocujú.



Obrázok 21. Model simulovanej siete v prostredí IT Guru

Výhodou prostredia OPNET IT Guru je možnosť vytvorenia viacerých profilov pre rôzne typy aplikácií, ich kombinácie a ich následné priradenia jednotlivým koncovým zariadeniam. V našom projekte sú klienti aplikácii HTTP a E-mail priradené zariadeniam stanica PC_1 – 5 a Notebook_1 – 4 a ich server zariadeniu HTTP_Email_Server. Počas priebehu simulácie budú teda koncové zariadenia generovať požiadavky vytvárajúce premávku v oboch smeroch, do aj zo siete. Nastavenia aplikácií môžeme meniť v bloku Definícia_Aplikácií, Obrázok 21.

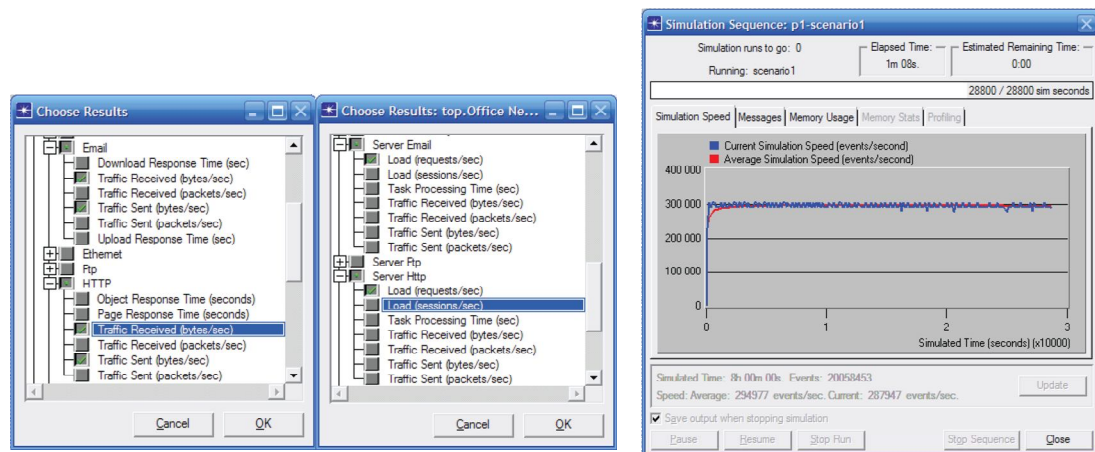


Obrázok 23. Definícia aplikácií s nastavením pravdepodobnosti prenosu.

Program umožňuje nastaviť atribúty pre oba zvolené druhy aplikácií s rôznymi parametrami ako veľkosť, početnosť a pravdepodobnosť výskytu. Zoznam typov rozdelenia pravdepodobnosti štandardne podporovanými programom OPNET IT Guru je Obrázku 23. vpravo. Snahou bolo nastaviť parametre tak, aby sa priblížili teoretickým poznatkom z kapitol 4 a 6 pre modelovanie HTTP a elektronickej pošty. Vytvorený projekt je uložený na priloženom optickom nosiči.

6.4 Simulácia modelu siete

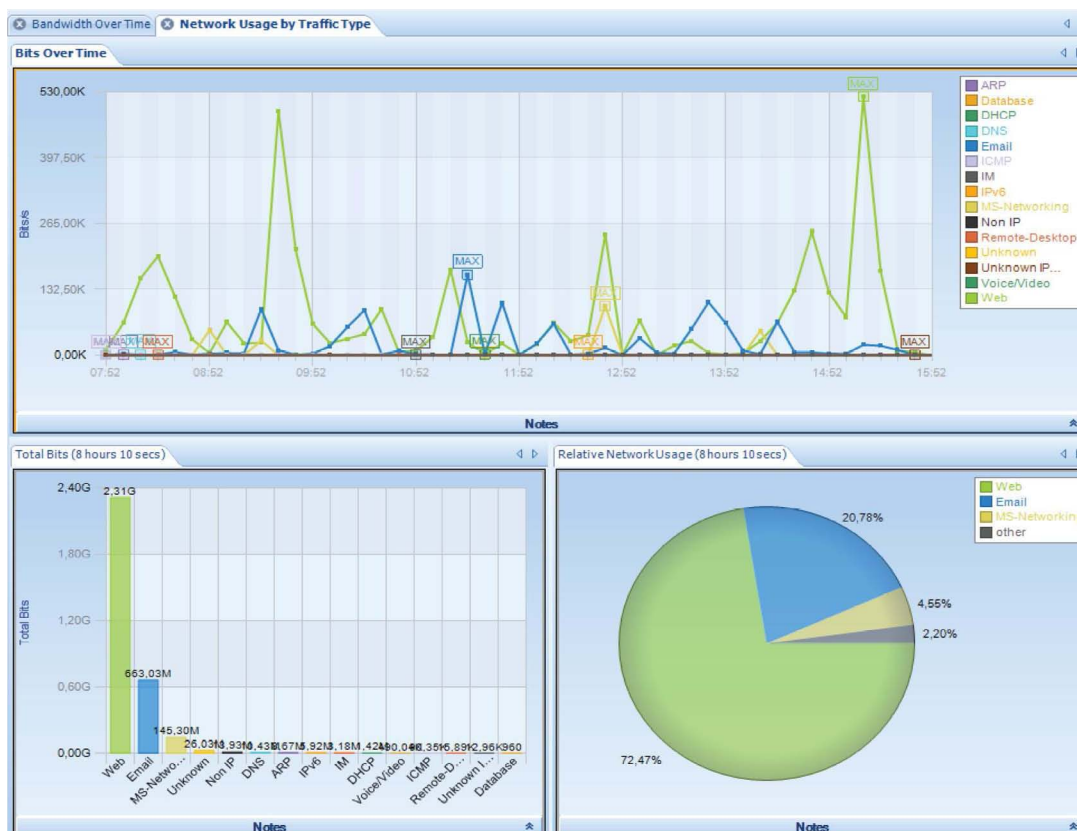
Po nastavení aplikačných parametrov je potrebné zvoliť štatistické údaje modelu siete, ktoré sú dôležité na výstupe simulácie. OPNET IT Guru umožňuje sledovať veľké množstvo štatistických údajov. Rozdeľujú sa na globálne alebo individuálne pre konkrétne zariadenie alebo uzol v sieti. Zo všetkých štatistík sú pre tento prípad najdôležitejšie súhrn požiadaviek HTTP a E-mail klientov v závislosti na čase simulácie (Requests/s). Túto štatistiku sledujeme na servery HTTP_E-mail_Server, ktorý v prípade simulácie reprezentuje súbor všetkých serverov, na ktoré bola odoslaná požiadavka request zo simulovanej siete(klienti HTTP a E-mail). Ďalej sledujeme globálne štatistiky vyslanej a prijatej premávky (Traffic received/sent) pre obe aplikácie v závislosti na čase simulácie. Po nastavení štatistických údajov podľa Obrázku 24 môžeme pristúpiť k spusteniu samotnej analýzy siete. Nastavíme čas analýzy $t = 8$ hodín a simulácia sa spustí tlačidlom „Run“. Simulácia úspešne končí po čase cca 1 min. Výstupné štatistiky sú dostupné vo forme grafov.



Obrázok 24. Voľba štatistík a výstupný graf simulácie.

6.5 Analýza prenosov v reálnej sieti

Na analýzu reálnej premávky bolo použité programové vybavenie Wireshark rozšírené o skúšobnú verziu CACE Pilot version 2.2, ktoré bolo nainštalované na stanici s označením Stat_PC. Stanica bola zapojená do siete LAN v uzle podľa Obrázku 25. Pomocou programu Wireshark bola odchytená všetka sieťová premávka v časovom intervale 8 hodín. Meranie prebehlo počas štandardného pracovného dňa. Meranie bolo úspešne realizované až po viacerých neúspešných pokusoch, v dôsledku vnútornej chyby programu Wireshark, keď sa podarilo sa zachytiť len krátky časový úsek, po ktorom program padol. Pôvodne predpokladané viacnásobné opakované merania tak nemohli byť vykonané. Podarilo sa však zaznamenať časový úsek 8 pracovných hodín, ktorý korešponduje s časom simulácie programu OPNET IT Guru. Súbor so zaznamenanými údajmi programu je umiestnený v priloženom optickom dátovom médiu pod názvom *test1.pcap*. Súbor je skomprimovaný a zašifrovaný heslom, ktoré má k dispozícii vedúci práce. Odchytené údaje boli ďalej podrobené analýze v programe CACE Pilot. Základnú informáciu o odchytenom súbore môžeme vidieť v Tabuľke 4. Podľa hodnôt výsledkov z grafov na Obrázku 25 je zrejmé, že najväčšiu zložku až 72,5% z celkového prenosu tvorí práve HTTP. Elektronická pošta má zastúpenie 21% s protokolom SNMP pre odoslanú a POP3 pre prijatú poštu, čo zodpovedalo očakávaniam spred realizácie odchyťovania premávky.



Obrázok 25. Zastúpenie protokolov v odchytenej premávke.

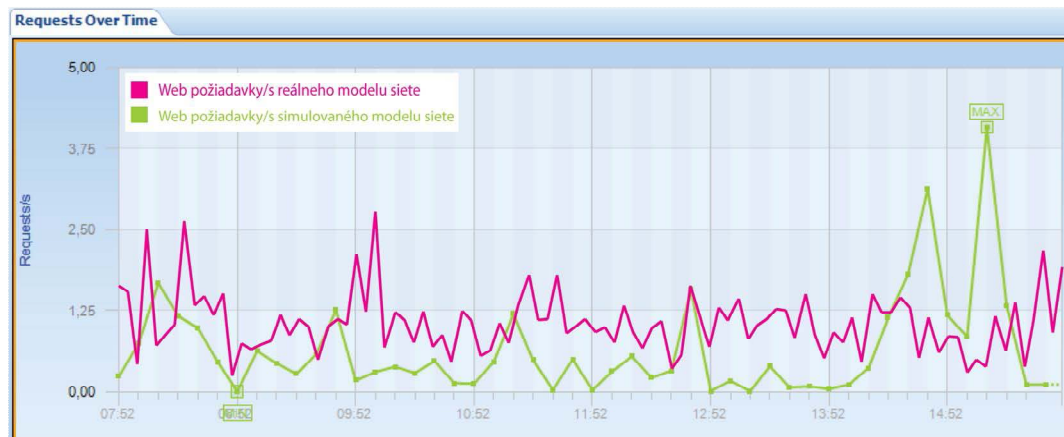
Statistic Name	Value
Total Number of Bits	3,191,457,544
Total Number of Bytes	398,932,193
Total Number of Packets	624,634
Number of Broadcast Packets	20,338
Number of Multicast Packets	30,846

Tabuľka 4. Štatistika súboru odchytenej premávky.

7.6 Simulovaná a reálna sieťová premávka

Základné vstupné parametre simulácie webovej premávky :

Verzia protokolu HTTP	HTTP 1.1
Čas prezerania stránky	10 s (konštatnta)
Veľkosť hlavného objektu	$LN(E = 10 \text{ kB}, \sigma = 25 \text{ kB})$
Počet vedľajších objektov	35 (konštatnta)
Veľkosť inline objektov	$LN(E = 7,7 \text{ kB}, \sigma = 126 \text{ kB})$
Veľkosť <i>getrequest</i>	360 B (konštatnta)



Obrázok 26. Grafický výstup zo simulovanej a reálnej webovej premávky

Základné vstupné parametre simulácie premávky elektronickej pošty :

Doba medzi dvoma vyslanými/prijatými E-mailami
Veľkosť E-mailu

E(180s)
LN(5 kB; 3,5 kB)



Obrázok 27. Grafický výstup zo simulovanej a reálnej premávky elektronickej pošty

Použité skúšobné verzie softvérových nástrojov a výkon výpočtovej techniky ponúkli iba obmedzené výstupy pre simuláciu a analýzu uvedených typov premávky. Uvedené grafické výstupy pre simulovanú a reálnu sieť poskytli iba obmedzený rozsah štatistických údajov, ktoré nepostačovali pre exaktné hodnotenie a vzájomné porovnanie oboch sietí. Ak by boli dostupné dostatočne rozsiahle štatistických súbory, bolo by možné porovnanie oboch sietí na základe štatistického vyhodnotenia údajov z oboch sietí. Ukázalo sa tiež, že v sieti malej spoločnosti sa značne prejavuje individuálne správanie užívateľov v priebehu pracovného dňa. V reálnej sieti sa nevyužívala VoIP telefónia a pokus o jej simuláciu na použitých prostriedkoch neposkytol použiteľné výstupy.

7. Záver

V práci je opísaný model sieťovej premávky podľa výskumu autorov uvedených v citovanej literatúre z pravdepodobnostného pohľadu. Tieto skúmania realizované v minulosti vychádzali zo štatistických údajov získaných zo sietí v súčasnosti pracujúcich s už minimálne využívanými spôsobmi prenosu, niektoré už vychádzali z novších technológií.

Dôležitým faktormi pri modelovaní sieťovej premávky je správanie užívateľov, spôsob akým využívajú konkrétne dátové služby, možnosti

poskytovaných technológií a forma poskytovaných služieb zo strany ich poskytovateľov. Na základe údajov v citovaných prácach je uvedená predikcia rozvoja jednotlivých služieb pri poskytovaní pripojenia cez ADSL, ktorá, ktorá síce dnes nie je najmodernejšia, ale je najrozšírenejšia v sektore domácich užívateľov, malých a stredných podnikov.

Práca sa zaoberá webovými prenosmi cez HTTP a dátovou telefóniou VoIP, ktoré v súčasnosti predstavujú najväčší podiel v dátových tokoch prebiehajúcich prenosov, čiastočne E-mailová premávka. Každá z týchto služieb má svoju vlastnú štruktúru a je modelovaná samostatne. Potrebné dáta, ktoré poskytli štatistický materiál a z nich získané pravdepodobnostné modely boli prevzaté z citovanej literatúry. Najdôveryhodnejšími údajmi z dátových a hlasových prenosov disponujú poskytovatelia dátových služieb a telekomunikačné spoločnosti, ktoré sa však od nich ťažko získavajú.

Model HTTP na základe správania sa užívateľa a činnosti protokolu je reprezentovaný rozdelením pravdepodobnosti lognormálnym pre počet požiadaviek na jednu reláciu, veľkosť hlavného objektu, veľkosti požiadavky *getrequest*, hlavné a vedľajšie objekty, Weibullovým pre doby prezerania stránok a gama pre počet inline objektov

Rozširovanie VoIP je v súčasnosti veľmi aktuálne a tvorí výrazný podiel z celkových dátových prenosov v sieti, preto bola v práci venovaná pozornosť aj modelovaniu sieťovej premávky práve pre túto technológiu.

V komunikácii VoIP sa pre samostatný model vydeľuje proces prichádzajúcich hovorov charakterizovaný Poissonovým rozdelením pravdepodobnosti a proces držania hovorov, ktorého pravdepodobnostný model je v zhode s Paretovým rozdelením pravdepodobnosti.

Skúmanie VoIP úrovni samotného prenosu hlasových paketov ukázalo, že kodeky detekciou hlasovej aktivity mali malý vplyv rozdelenia periód ON/OFF, na ktoré je vhodné aplikovať agregačné modely zodpovedajúce rozdeleniu s ťažkým chvostom, konkrétne Paretovým. Naznačuje to, že agregované dátové toky sa vyznačujú vnútornou podobnosťou.

Zisťovanie pravdepodobnostného modelu pre elektronickú poštu predstavuje najčastejšie vyhodnocovanie veľkosti prenášaných E-mailov. Dobré priblíženie nameraných údajov, najmä pre veľké E-maily poskytuje Cauchyho rozdelenie pravdepodobnosti.

Z poznatkov uvedených v práci vyplýva, že každá zo služieb má svoj charakteristický pravdepodobnostný model, ktorý je určený spôsobom akým je používaná, akou technológiou a na akých protokoloch aplikácia prenáša údaje, akými mechanizmami ovplyvňujú sieťové uzly siete prenosové parametre, v ktorých miestach sa sleduje premávka a zrejme ďalšími faktormi.

Znamená to, že aplikovať model jednej sieťovej služby nie je možné na službu, ktorá je zo svojej podstaty odlišná. Výsledný univerzálny model by bolo možné vytvoriť kompozíciou uvedených modelov pre jednotlivé služby, čo by predstavovalo náročnú matematickú úlohu a výsledkom by bol pravdepodobne viac - parametrický pravdepodobnostný model sieťovej premávky.

Súčasťou práce bola realizácia softvérovej simulácie, pre ktorú bol použitý programový nástroj OPNET IT Guru Academic Edition 9.1, ktorý po vytvorení projektu modelu reálnej siete a nastavení parametrov premávky pre webové prenosy a elektronickú poštu umožnil nasimulovať jeden 8 - hodinový pracovný deň malej spoločnosti. Pri nastavení simulácie bolo snahou využiť v práci uvedené teoretické poznatky zo skúmaní, ktoré boli vykonané v minulosti autormi v citovanej literatúre.

Doplnením softvérovej simulácie bolo odchytenie a analýza sieťovej premávky v rozsahu 8-hodinového pracovného dňa malej reklamnej spoločnosti pomocou programového nástroja WireShark rozšíreného o skúšobnú verziu CACE Pilot version 2.2.

Možnosti simulačného nástroja v skúšobnej verzii poskytli iba obmedzené údaje, ktoré umožnili iba kvalitatívne porovnanie simulovanej a reálnej siete. Pri predpoklade použitia programu v plnej verzii na dostatočne výkonnom výpočtovom prostriedku by možné simulovať rozsiahlejšie siete a poskytnúť užitočné údaje pre hodnotenie prevádzky či už pripravovaných alebo existujúcich sieťových riešení.

8. Dodatok

8.1 Vybrané pojmy z teórie pravdepodobnosti a štatistiky

Cieľom práce je vytvorenie modelu sieťovej prevádzky, ktorý využíva je postavený na teórii pravdepodobnosti a teórii hromadnej obsluhy. V tejto kapitole sú uvedené základné definície pojmov predovšetkým z [16], [17] a internetových zdrojov, predovšetkým z encyklopédie wikipédia.

Definícia 8.1

Náhodná premenná X je funkcia, ktorá každému možnému výsledku priradí hodnotu z množiny reálnych čísel. Náhodná premenná môže byť diskrétna, keď nadobúda hodnoty z konečnej alebo spočítateľnej množiny, alebo spojitá, keď môže nadobúdať ľubovoľnú hodnotu z intervalu reálnych čísel.

Definícia 10.2

Distribučná funkcia (cdf) náhodnej premennej X je pravdepodobnosť, že náhodná veličina X má hodnoty menšie alebo rovné reálnemu číslu x :

$$F(x) = P(X \leq x)$$

Komplementárna distribučná funkcia (ccdf) je definovaná :

$$F_c(x) = 1 - F(x)$$

Definícia 8.3

Funkcia $f(x)$ sa nazýva hustota pravdepodobnosti (pdf) spojitej náhodnej premennej X , ak pre ľubovoľné reálne číslo $x_0 \in (-\infty, \infty)$:

$$F(x_0) = \int_{-\infty}^{x_0} f(x) dx$$

$$F(x) = \int_{-\infty}^x f(t) dt$$

$$f(x) = \frac{d}{dx} F(x)$$

Definícia 8.4

Stredná hodnota diskretnej náhodnej premennej X :

$$E(x) = \sum_{i=1}^n x_i p(x_i)$$

Stredná hodnota spojitaj náhodnej premennej X :

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

Definícia 8.5

Disperziou náhodnej premennej X nazývame číslo :

$$D(x) = E[(X - E(x))^2]$$

Smerodajná odchýlka premennej X nazývame číslo :

$$\sigma(x) = \sqrt{D(x)}$$

Definícia 8.6

Náhodná premenná X má binomické rozdelenie $Bi(n,p)$ pravdepodobnosti s parametrami n ; p práve vtedy, ak pravdepodobnosť, že pri n nezávislých pokusoch nastane pozorovaný jav práve k -krát :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad i = 0, 1, 2, \dots, n$$

Stredná hodnota: $E(X) = np$

Disperzia: $D(X) = np(1-p)$

Definícia 8.7

Náhodná premenná X má exponenciálne rozdelenie pravdepodobnosti $\text{Exp}(\lambda)$ s parametrom $\lambda \in R^+$ práve vtedy, ak hustota pravdepodobnosti má tvar:

$$f(x) = \lambda e^{-\lambda x} \quad x > 0, \lambda > 0$$

Stredná hodnota: $E(X) = \lambda^{-1}$

Disperzia: $D(X) = \lambda^{-2}$

Definícia 8.8

Náhodná premenná X má Poissonovo rozdelenie pravdepodobnosti $Po(\lambda)$ s parametrom $\lambda \in R^+$ práve vtedy, ak pravdepodobnostná funkcia má tvar:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad i = 0, 1, 2, \dots, n$$

Stredná hodnota: $E(X) = \lambda$

Disperzia: $D(X) = \lambda$

Poissonovo rozdelenie je limitný prípad binomického rozdelenia $n \rightarrow \infty$ a $p \rightarrow 0$. V praxi sa toto rozdelenie používa na aproximáciu málo pravdepodobných náhodných udalostí pri veľkom počte nezávislých opakovaní experimentu.

Definícia 8.9

Náhodná premenná X má Erlangovo rozdelenie pravdepodobnosti $E(n, \lambda)$ práve vtedy, ak hustota pravdepodobnosti má tvar:

$$f(x) = \frac{\lambda^{n+1}}{n!} e^{-\lambda x} x^n \quad \lambda > 0, x > 0, n = 0, 1, \dots$$

Stredná hodnota: $E(X) = n / \lambda$

Disperzia: $D(X) = n / \lambda^2$

Pre $n = 0$ prechádza Erlangovo rozdelenie na exponenciálne.

Definícia 8.10

Náhodná premenná X má normálne rozdelenie pravdepodobnosti $N(\mu, \sigma)$ s parametrami μ a σ^2 práve vtedy, ak hustota pravdepodobnosti má tvar:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in (-\infty, \infty), \mu \in (-\infty, \infty), \sigma > 0$$

Stredná hodnota: $E(X) = \mu$

Disperzia: $D(X) = \sigma^2$

V praxi môžeme Binomické rozdelenie $Bi(n, p)$ aproximovať normálnym, ak $np > 5$ a $n(1-p) > 5$ s parametrami np a $np(1-p)$ a Poissonovo rozdelenie $Po(\lambda)$, ak $\lambda > 5$ s parametrami λ a λ .

Definícia 8.11

Náhodná premenná X má lognormálne rozdelenie pravdepodobnosti $LN(\mu, \sigma)$ s parametrami μ a σ^2 práve vtedy, ak $\ln(X)$ má normálne rozdelenie pravdepodobnosti s parametrami μ a σ^2 a hustota pravdepodobnosti má tvar:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad x \in (0, \infty), \mu > 0, \sigma > 0$$

Stredná hodnota: $E(X) = e^{\mu + \frac{\sigma^2}{2}}$

Disperzia: $D(X) = (e^{\sigma^2} - 1) e^{2\mu + \sigma^2}$

Definícia 8.12

Náhodná premenná X má gama rozdelenie pravdepodobnosti $G(\alpha, \beta)$ s parametrami α, β práve vtedy, ak hustota pravdepodobnosti má tvar :

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} \quad x > 0, \alpha > 0, \Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Stredná hodnota: $E(X) = \alpha/\beta$

Disperzia: $D(X) = \alpha/\beta^2$

Ak parameter α je celé číslo, potom gama rozdelenie reprezentuje súčet α nezávislých náhodných premenných s exponenciálnym rozdelením s parametrom β . Ak $\alpha = 1$, gama rozdelenie $G(1, \beta)$ je totožné s exponenciálnym rozdelením $\text{Exp}(\beta)$.

Definícia 8.13

Náhodná premenná X má Paretovo rozdelenie pravdepodobnosti $\text{Pa}(\alpha, b)$ s parametrami α, b práve vtedy, ak hustota pravdepodobnosti má tvar :

$$f(x) = \frac{\alpha b^\alpha}{(b+x)^{\alpha+1}} \quad x > 0, \alpha > 0, b > 0$$

Stredná hodnota: $E(X) = b/(\alpha-1), \alpha > 1$

Disperzia: $D(X) = \alpha b^2 / [(\alpha-1)^2 (\alpha-2)], \alpha > 2$

Definícia 8.14

Náhodná premenná X má všeobecné Paretovo rozdelenie pravdepodobnosti (GDP) $GPa(\alpha, b, \mu)$ s parametrami α (tvar), b (mierka), μ (prahová hodnota X) práve vtedy, ak hustota pravdepodobnosti má tvar :

$$f(x) = \frac{1}{b} \left(1 + \frac{\alpha(x - \mu)}{b}\right)^{-\frac{1}{\alpha}-1} \quad x \geq \mu, \alpha \geq 0; \mu < x \leq \mu - b/\alpha, \alpha < 0$$

Stredná hodnota: $E(X) = \mu - b/(\alpha-1), \alpha < 1$

Disperzia: $D(X) = b^2/[(1-b)^2(1-2b)], \alpha < 1/2$

Definícia 8.15

Náhodná premenná X má Weibullovo rozdelenie pravdepodobnosti $W(\lambda, k)$ s parametrami λ, k práve vtedy, ak hustota pravdepodobnosti má tvar :

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \quad x > 0, k > 0, \lambda > 0$$

Stredná hodnota: $E(X) = \lambda \Gamma(1+1/k)$

Disperzia: $D(X) = \lambda^2 [\Gamma(1+2/k) - \Gamma(1+1/k)^2]$

Definícia 8.16

Náhodná premenná X , ktorá je súčtom štvorcov n nezávislých veličín s normovaným normálnym rozdelením $N(0,1)$, má χ^2 rozdelenie pravdepodobnosti $Chi(n)$ s parametrom n (počet stupňov voľnosti) práve vtedy, ak hustota pravdepodobnosti má tvar :

$$f(x) = \frac{1}{\sqrt{2^n} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad x > 0, n \in N$$

Stredná hodnota: $E(X) = n$

Disperzia: $D(X) = 2n$

Definícia 8.17

Náhodná premenná X má Cauchyho rozdelenie pravdepodobnosti $C(\alpha, \lambda)$ s parametrami α (najväčšia hodnota rozdelenia), λ (mierka) práve vtedy, ak hustota pravdepodobnosti má tvar :

$$f(x) = \left\{ \pi \lambda \left[1 + \left(\frac{x - \alpha}{\lambda} \right)^2 \right] \right\}^{-1} \quad \lambda > 0, -\infty < \alpha \leq \infty$$

Stredná hodnota: nemá
Disperzia: nemá

Zoznam použitej literatúry

- [1] CHOI, H. - LIMB, J. : A Behavioral Model of Web Traffic, International Conference of Networking Protocol 99' (ICNP 99'), september 1999.
- [2] DANG,T.D.- SONKOLY, B., - S. MOLNAR, S. : Fractal analysis and modeling of VoIP traffic, NETWORKS 2004, s.123–130, Vienna, Austria, jún 2004.
- [3] KLEMM, A. - LINDEMANN, A. - LOHMANN, M. : Traffic Modeling and Characterization for UMTS Networks, Proc. of the Globecom, Internet Performance Symposium, San Antonio, s. 1741-1746, 2001.
- [4] STAEHLE, D.- LEIBNITZ, K.-TRAN-GIA, P. : Source Traffic Modeling of Wireless Applications, Universität Würzburg, Institut für Informatik, Würzburg, 2000
- [5] MAH, B. : An Empirical Model of HTTP Network Traffic, Proceedings of the IEEE INFOCOM 97, Kobe, apríl 1997, vol. 2
- [6] MEIER-HELLSTERN, K. - WIRTH, P.- Yan, Y.L.- Hoeflin, D. A. : Traffic Models for ISDN Data Users: Office Automation Application, Proc. Of International Teletraffic Congress (ITC-13), A. Jensen and V. B. Iversen (Eds.), Elsevier, North-Holland, 1991.
- [7] PAPOULIS, A. - PILLAJ,S.U. : Probabilities, Random Variables and Stochastic Process, McGraw-Hill, New York 2002.
- [8] PAXSON, V. : Empirically-derived analytic models of widearea tcp connections, IEEE/ACM Transactions on Networking, 1994.
- [9] PAXSON, V. - Floyd, S. : Wide Area Traffic: The Failure of Poisson Modeling, IEEE/ACM Transactions on Networking, vol. 3, no. 3, s. 226-244, 1995.
- [10] SCHULZRINNE, H. : RTP Tools,
<http://www.cs.columbia.edu/IRT/software/rtptools/>
- [11] SELSIUS Inc., 11 Client Control Protocol.
- [12] VICARI, N. : Measurement and Modeling of WWW-Sessions, Institute of Computer Science, University of Würzburg, Technical Report No. 184
- [13] VICARI, N. - KOEHLER, S. : Measuring Internet User Traffic Behavior Dependent on Access Speed, Institute of Computer Science, University of Würzburg, Technical Report No. 238, október 1999.

- [14] N. Provos, N. : Voice over misconfigured internet telephones,
<http://14.xtdnet.nl/>.
- [15] DERI, L.: Open Source VoIP Traffic Monitoring, Proceedings of SANE 2006, May 2006
- [16] LAMMOŠ, F.- POTOCKÝ, R.: Pravdepodobnosť a matematická štatistika, Bratislava, Vydavateľstvo UK, 1998, 344 s., ISBN 80-223-1262-2.
- [17] J. VENTCEL'OVÁ, J.: Teória pravdepodobnosti, Bratislava, Vydavateľstvo Alfa, 1973, 524 s., ISBN 65-553-73
- [18] REYES-LECUONA, A., GONZÁLES-PARADA, E., CASILARI, E., CASSSAOLA, J.C., DIAZ-ESTRELLA, A.: A page-oriented WWW traffic model fo wireless system simulations, Proceedings of 16th ITC, Edinburgh, June, 1999
- [20] Lourens O. W.: A Web browsing workload model for simulation, dissertation, department of computer science, Faculty of Science at the University of Cape Town, May 2004

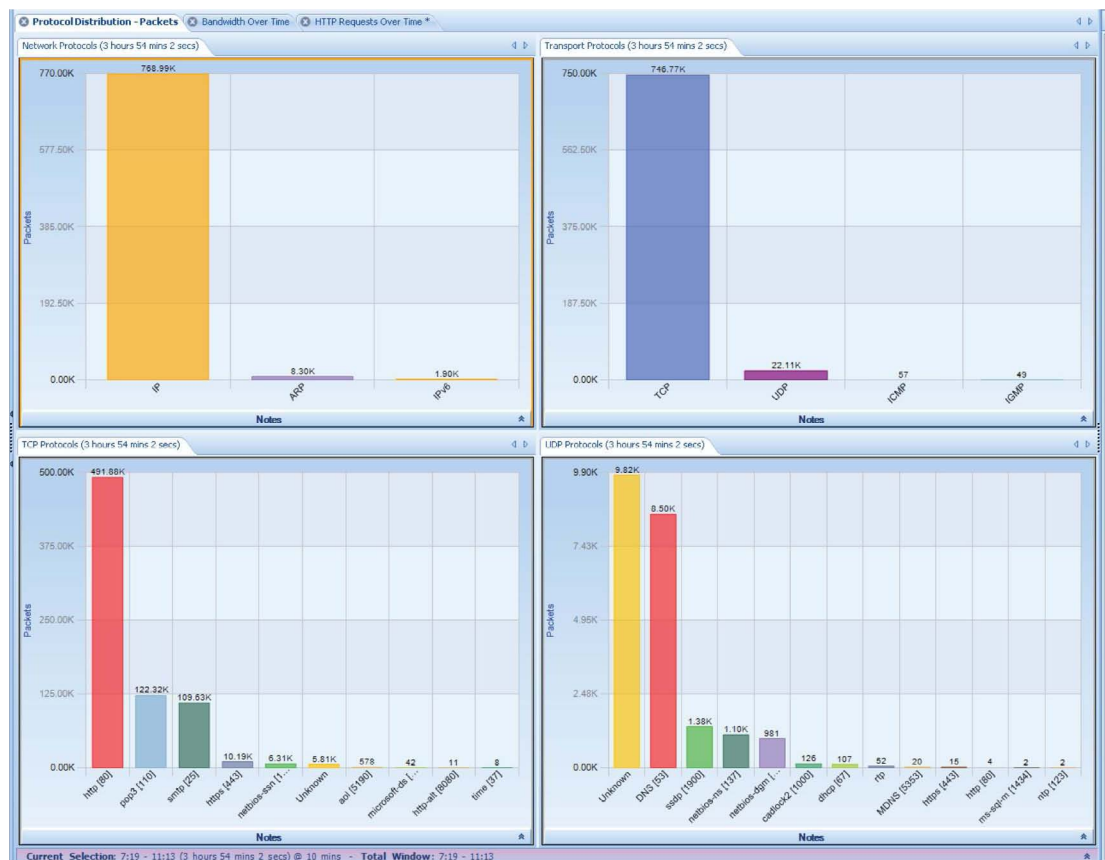
Zoznam použitých skratiek

ASCII	<i>American Standard Code for Information Interchange</i>
ATM	<i>Asynchronous Transfer Mode</i>
DNS	<i>Domain Name System</i>
DSL	<i>Digital Subscriber Line</i>
DSLAM	<i>Digital Subscriber Line Access Multiplexer</i>
FTP	<i>File transport Protocol</i>
GPD	<i>Generalized 7o Distribution</i>
GPRS	<i>General Packet Radio Services</i>
GSM	<i>Global System for Mobile Communications</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IP	<i>Internet Protocol</i>
ISP	<i>Internet Service Provider</i>
LAN	<i>Local Area Network</i>
LDR	<i>Long Range Dependence</i>
NNTP	<i>Network News Transfer Protocol</i>
P2P	<i>Peer – to - Peer</i>
PBX	<i>Private Branch Exchange</i>
PC	<i>Personal Computer</i>
POP3	<i>Post Office Protocol</i>
PSTN	<i>Public Switched Telephone Network</i>
QoS	<i>Quality of Services</i>
RTCP	<i>Real-time Transport Control Protocol</i>
RTCP XS	<i>Real-time Transport Control Protocol Extended</i>
RTP	<i>Real-time Transport Protocol</i>
SIP	<i>Session Initiation protocol</i>
SMTP	<i>Simple Mail Transfer Protocol</i>
STM	<i>Synchronous Transfer Mode</i>
TCP	<i>Transmission Control Protocol</i>
UDP	<i>User Datagram Protocol</i>
URL	<i>Uniform Resource Locator</i>
VoIP	<i>Voice over Internet Protocol</i>
WAN	<i>Wide Area Network</i>
Wi-Fi	<i>Wireless Fidelity</i>
WWW	<i>World Wide Web</i>

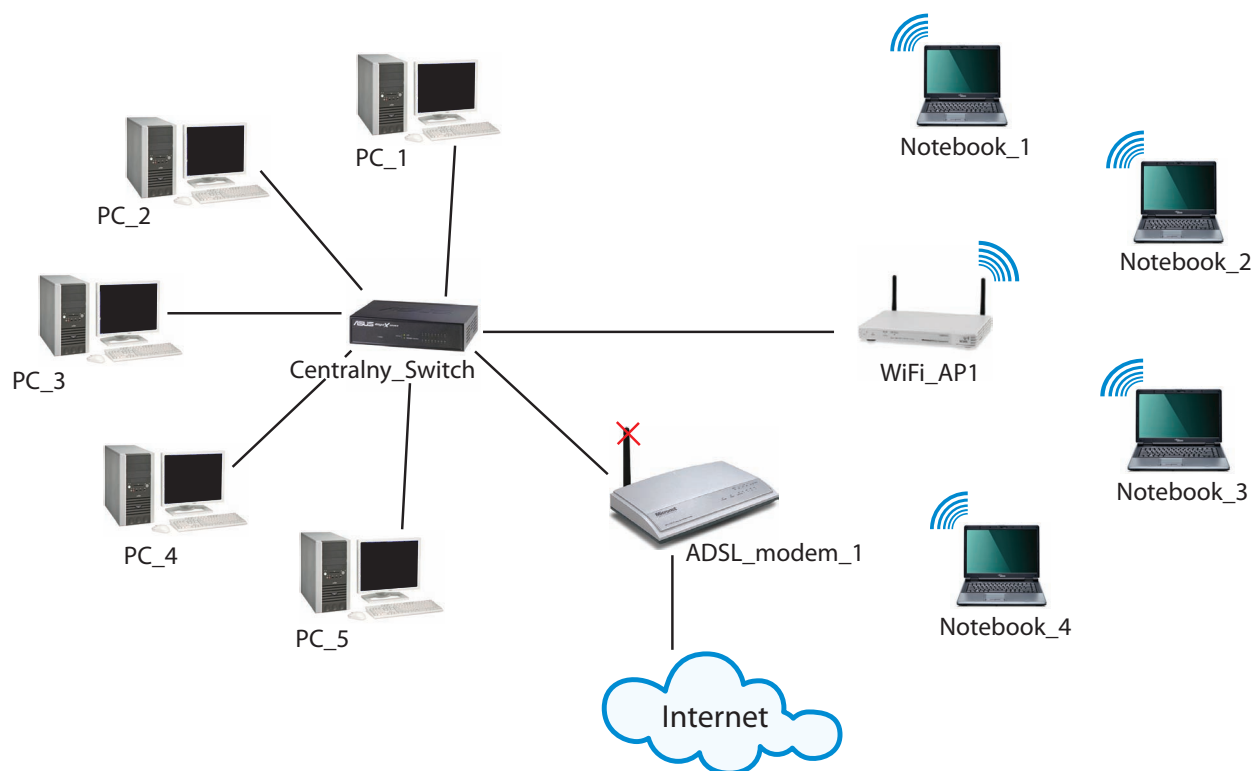
Zoznam príloh

1. CD s obsahom elektronickej formy záverečnej práce
2. Súbor simulovaného referenčného modelu HTTP a E-mail premávky siete v prostredí IT Guru Academic edition uložené na priloženom CD
3. Súbor odchytených dát prenosu analyzovanej siete programu WireShark.
4. Obrázok 15. Topológia reálnej siete
Obrázok 16. Zapojenie pre odchytyvanie paketov
Obrázok 17. Prenosy aplikačných protokolov HTTP, SMTP a POP3
Obrázok 20. Scenár projektu modelu siete OPNET IT Guru

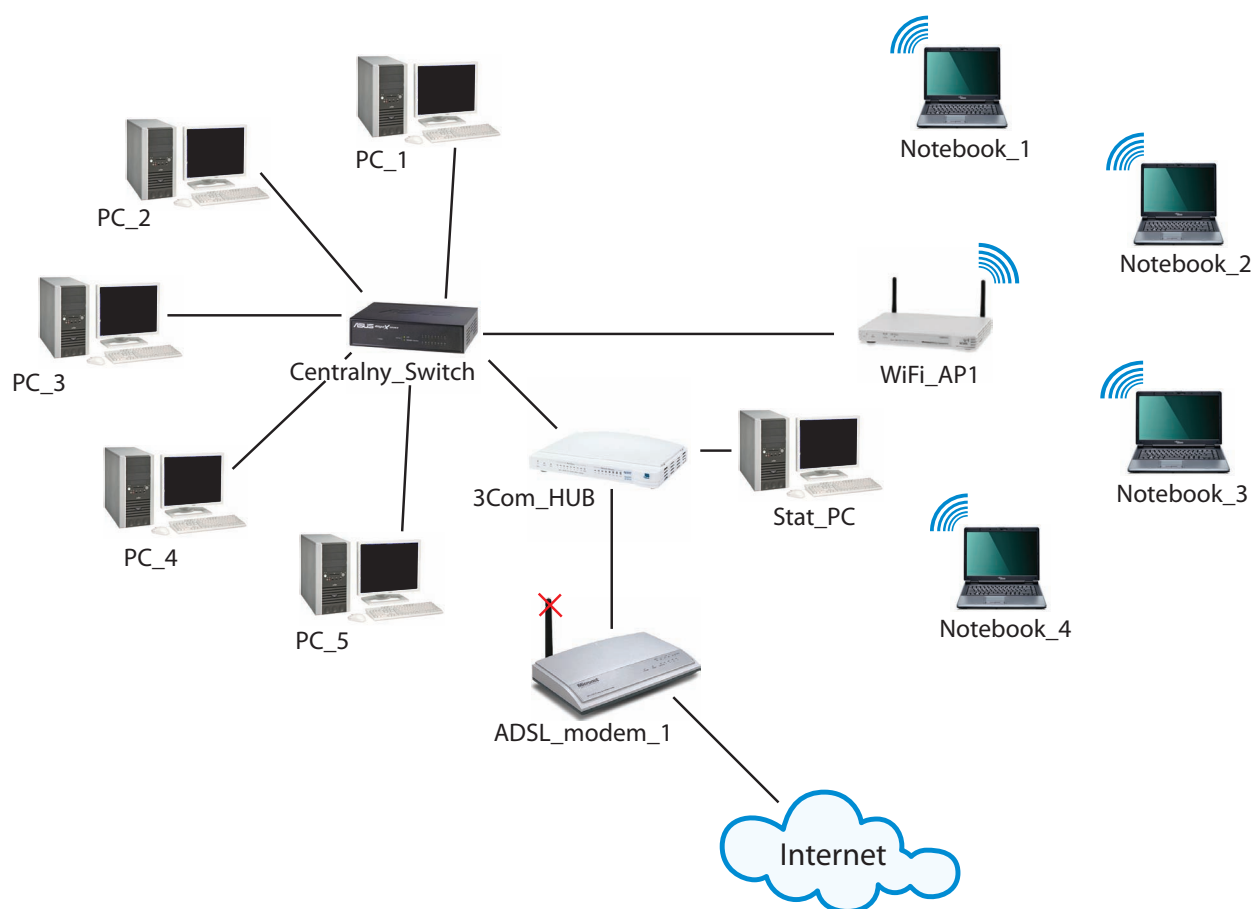
Prílohy



Obrázok 17. Prenosy aplikačných protokolov HTTP, SMTP a POP3

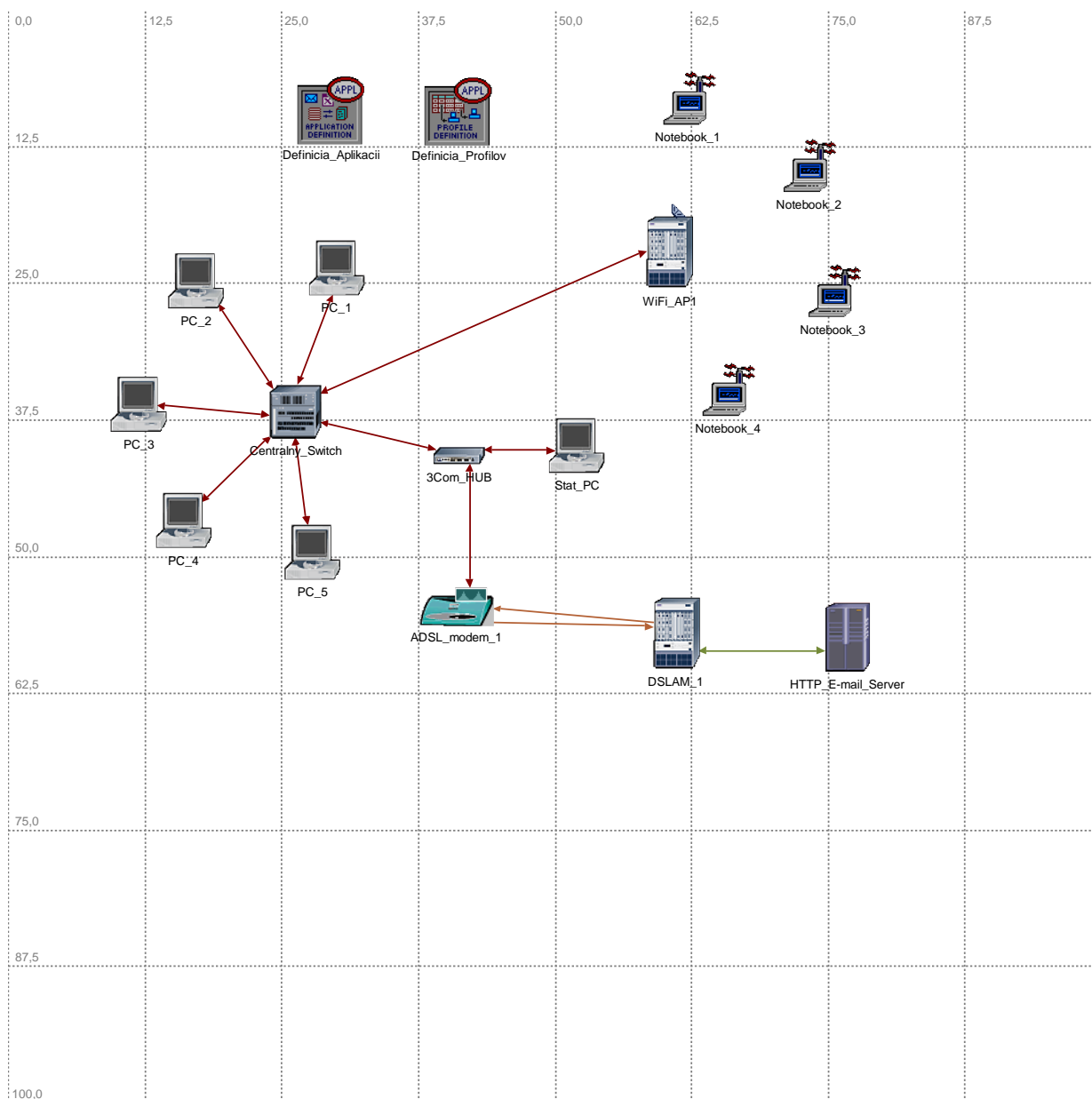


Obrázok 15. Topológia reálnej siete



Obrázok 16. Zapojenie pre odchytyvanie paketov

Network: Scenario1 [Subnet: top.Office Network]



Obrázok 20. Scenár projektu modelu siete OPNET IT Guru